Learning to Interpret a Disjunction

Masoud Jasbi[1], Akshay Jaggi[2], & Michael C. Frank[2]

[1] Harvard University

[2] Stanford University

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

Correspondence concerning this article should be addressed to Masoud Jasbi, Postal address. E-mail: masoud_jasbi@fas.harvard.edu

Abstract

At first glance, children's word learning appears to be mostly a problem of learning words like *dog* and *run*. However, it is small words like *and* and *or* that enable the construction of complex combinatorial language. How do children learn the meaning of these function words? Using transcripts of parent-child interactions, we investigate the cues in child-directed speech that can inform the interpretation and acquisition of the connective *or* which has a particularly challenging semantics. Study 1 finds that, despite its low overall frequency, children can use *or* close to parents' rate by age 4, in some speech acts. Study 2 uses annotations of a subset of parent-child interactions to show that disjunctions in child-directed speech are accompanied by reliable cues to the correct interpretation (exclusive vs. inclusive). We present a decision-tree model that learns from a handful of annotated examples to correctly predict the interpretation of a disjunction. These studies suggest that conceptual and prosodic cues in child-directed speech can provide information for the acquisition of functional categories like disjunction.

Keywords: keywords

Word count: X

Learning to Interpret a Disjunction

Introduction

Word learning is commonly construed as the process of isolating a word form, selecting a meaning from a set of potential meanings, and mapping the word to the selected meaning (**???**). For example, a father holding a baby may point to a squirrel and say "look at the squirrel!" The baby – already familiar with the phrase "look at the" – should recognize the novel word *squirrel*, consider some potential referents (e.g tree, squirrel, chair, etc.) and select the right referent using the available cues, in this case the father's pointing. While there has been a lot of research on cues and mechanisms that help children's acquisition of content words such as *squirrel*, *red*, and *run*, cues and mechanisms that can assist children in learning the meaning of function words such as *and* and *or* have remained a major challenge. In this study, we focus on the disjunction word *or* and provide a novel learning account that uses salient cues to learn the interpretations of disjunction in English.

How do children learn the meaning of disjunction

Previous Studies

To our knowledge, only one study has looked at spontaneous productions of *and* and *or* in parents' and children's speech. Morris (2008) investigated children between the ages of 2;0 and 5;0, using 240 transcriptions of audiotaped exchanges obtained from the CHILDES database. Each connective was analyzed with respect to its frequency, sentence type, and meaning (or use). The study found that overall, *and* was approximately 12.8 times more likely to be produced than *or*. The connective *and* appeared predominantly in statements (more than 90% of the time) while *or* was most common in questions (more than 85% of the time). Children started producing *and* at 2 years and *or* at 2.5 years of age.

Regarding the meaning of the connectives, Morris (2008) adopted a usage-based (item-based) approach (Levy & Nelson, 1994; Tomasello, 2003) and predicted that children start producing connectives with a single "core meaning" (also referred to as "use" or "communicative function"). He predicted that the core meaning mirrors the most frequent usage/meaning of the connective in child-directed speech. Children acquire the less frequent meanings of the connectives as they grow older. He found that children started producing *and* as conjunction at 2, and *or* as exclusive disjunction at 2.5 years of age. In line with the predictions of the usage-based account, he found that these two meanings are the most frequent meanings in parents' speech. For disjunction, 75-80% of the *or*-examples children heard recevied an exclusive interpretation. Finally, as children grew older, they started using connectives to convey additional meanings such as inclusive disjunction for *or* and temporal conjunction for *and*. However, the inclusive use of *or* was extremely rare in adults, and children barely produced it even at age 5. Morris (2008) argued that the development of connectives conforms to the predictions of a usage-based account and that in the first five years of children's development, the (core) meaning of disjunction is exclusive.

However, a series of experimental studies have found that preschool children are more likely to interpret *or* as inclusive in a variety of linguistic contexts such as negative sentences (Crain, Gualmini, & Meroni, 2000), conditional sentences (Gualmini, Crain, & Meroni, 2000), restriction and nuclear scope of the universal quantifier *every* (Chierchia, Crain, Guasti, Gualmini, & Meroni, 2001; Chierchia et al., 2004), nuclear scope of the negative quantifier *none* (Gualmini & Crain, 2002), restriction and nuclear scope of *not every* (Notley et al., 2012a), and prepositional phrases headed by *before* (Notley et al., 2012b). These studies almost unanimously claim that at least in declarative sentences, the inclusive interpretation of *or* emerges earlier than the exclusive interpretation.

The findings of these studies as well as those of Morris (2008) give rise to what we call "the paradox of learning disjunction". Given Morris (2008)'s finding that the majority of *or* examples children hear are exclusive, how can children learn to interpret it as inclusive? One way to addresses this paradox is logical nativism [crain2008logic; crain2010logic;@crain2012emergence]. This approach assumes that the language faculty contains information regarding what connective meanings are allowed for connective words crosslinguistically. Crain (2012) considered it unlikely that children learn the meaning of *or* from the examples they hear in adult usage. Instead, he argued that children rely on an innate knowledge that the meaning of disjunction words in natural languages must be inclusive. In other words, upon hearing a connective word, children consider inclusive disjunction as a viable candidate for its meaning but not exclusive disjunction. In this account, the exclusive interpretation emerges as part of children's pragmatic development after they have mastered the inclusive semantics of disjunction.

While logical nativism addresses the paradox of learning disjunction, it does not provide an explanation for cases where children interpret disjunction as exclusive. Morris (2008) reported that in his study, the vast majority of children used *or* in its exclusive sense. This is not expected if preschool children consider disjunction to be inclusive. Second, other experimental studies, especially those testing disjunction in commands, find that preschool children interpret it as exclusive (Braine & Rumain, 1981; Johansson & Sjolin, 1975). For example, in response to a command such as "give me the doll or the dog", children as young as three- and four-years-old give one of the objects and not both. In its current version, the nativist account does not provide any explanation for such cases.

Figure 1 summarizes the usage-based and nativist approaches to the acquisition of disjunction. The major difference between them is their assumptions on the learners' semantic hypothesis space for *or*. The usage-based account considers a wide array of meanings to be available for mapping, including different flavors of conjunction such as "temporal conjunction" (e.g. Bob pressed the key and (then) the door opened) and "explanatory conjunction". The nativist account limits the hypothesis space to binary logical

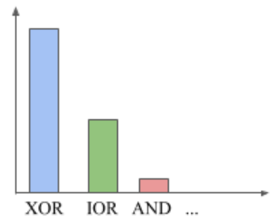
Learning Accounts of Disjunction	Binary Connective Hypothesis Space	Input Frequency for *or*	Early Mapping
Usage-Based Account (Morris 2008)	{XOR, IOR, IF, AND, $\text{AND}_{\text{temporal}}$, $\text{AND}_{\text{explanatory}}$, $\text{AND}_{\text{extension}}$, ...}		"or" → XOR
Logical Nativism (Crain 2012)	{IOR, AND, IF}		"or" → IOR

Figure 1. Summary of the usage-based and nativist approaches to the acquisition of disjunction.

connectives, more specifically to those commonly used in standard propositional logic: inclusive disjunction, conjunction, and material implication. Both accounts agree that the input favors the exclusive interpretation of disjunction. The usage-based account concludes that children's early mappings mirror this input. The nativist account suggests that innate biases towards the inclusive meaning and against the exclusive interpretation result in an inclusive semantics for *or* in children's early mappings.

Current Study

In this study, we provide an alternative solution to the paradox of learning disjunction. The main claim of this paper is that children may learn to interpret *or*– for example as exclusive or inclusive – using the salient cues that accompany it in the input. We support this hypothesis using three studies. In the first study, we investigate the distribution of *and* and *or* in parent-child interactions to address the following basic questions: how often do children hear or produce *or*? when do they start producing it? Using a large corpus of parent-child interactions, we found that children hear 1-2 examples of *or* in every thousand words parents produce. They start producing it themselves between 18-30 months, and by 42 months they reach a rate of one *or* per thousand words. In study 2, we ask: what interpretations can *or* have in child-directed speech? We annotated examples of *or* and found that its most likely interpretation is exclusive disjunction, as Morris (2008) had concluded. However, we also found that exclusive interpretations correlated strongly with two cues: rise-fall prosody, and logically inconsistent propositions connected by *or*. In the absence of these cues, *or* was most likely inclusive. In our third study, we ask if it is possible to learn the interpreation of *or* from these cues. Using the annotation data of study 2 and a supervised learning task, we showed that a decision-tree classifer can use prosody and consistency of propositions to predict its interpretation with high accuracy.

Based on the results of our studies, we propose a new account for children's acquisition of disjunction. Figure 2 shows the summary of this account which we call cue-based context-dependent mapping. It is inspired by the usage-based and nativist accounts of disjunction and shares many of their insights. Similar to the nativist account, we assume that the semantic hypothesis space includes binary logical relations. However, we do not limit the hypothesis space further and do not bias the learning towards the inclusive meaning. We will show that the input can achieve this. Similar to usage based proposals, our account relies on the structure of the input to distinguish between exclusive and inclusive uses of disjunction. We also map more complex constructions to meanings rather than the word *or* directly. The learner can later extract commonalities across these mappings and extract a core semantics for a particular word. However, the early mappings do not have any core meaning as opposed to what the usage-based account of Morris (2008) proposes. The major point of departure from previous accounts is the mechanism of learning. While in pervious accounts the most frequent meaning in the input was mapped to the connective word directly, in our account the input is partitioned or broken down by a set of salient cues that designate the context of use. Mapping is done based on the cues that accompany the connective word.

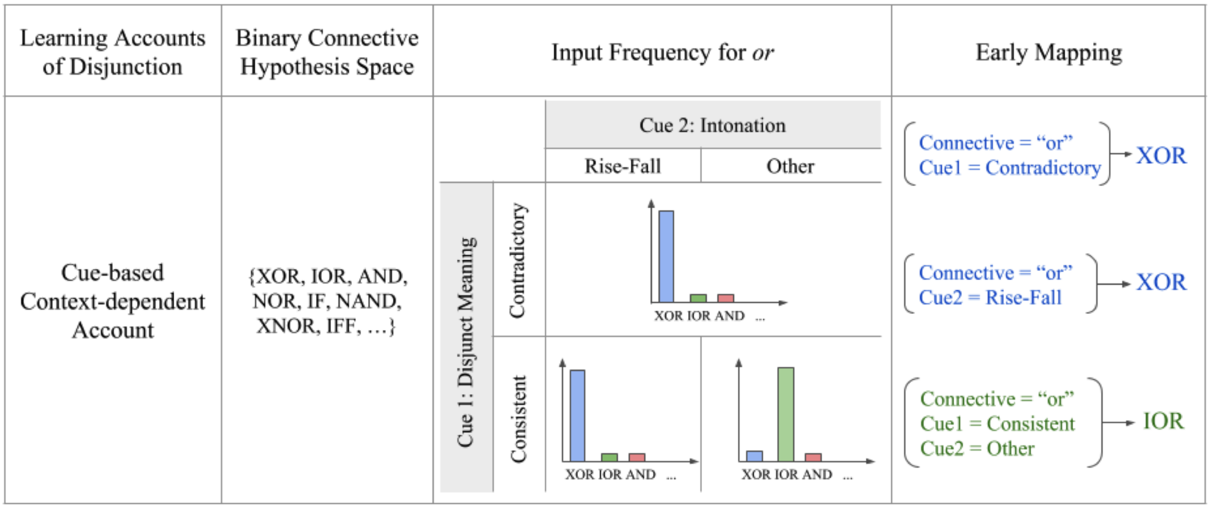


Figure 2. Summary of the usage-based and nativist approaches to the acquisition of disjunction.

Study 1: Production of "or" in parent-child interactions

In our first study, we looked at the frequencies of *and* and *or* in a corpus of parent-child interactions (CHILDES) with 14,159,609 words. This is a considerably larger corpus than previously used.

Methods

For samples of parents' and children's speech, we used the online database childes-db and its associated R programming package `childesr` (Sanchez et al., 2018). Childes-db is an online interface to the child language components of TalkBank, namely CHILDES

(MacWhinney, 2000) and PhonBank. Two collections of corpora were selected: English-North America and English-UK. All word tokens were tagged for the following information: 1. The speaker role (mother, father, child), 2. the age of the child when the word was produced, 3. the type of the utterance the word appeared in (declarative, question, imperative, other), and 4. whether the word was *and*, *or*, or neither.

Exclusion Criteria. First, tokens were coded as unintelligible were excluded (N = 290,119). Second, tokens that had missing information on children's age were excluded (N = 1,042,478). Third, tokens outside the age range of 1 to 6 years were excluded (N = 686,870). We were interested in the 1 to 6 years old age range and there was not much data outside this age range. The collection contained the speech of 504 children and their parents after the exclusions.

Procedure. Each token was marked for the utterance type that the token appeared in. This study grouped utterance types into four main categories: "declarative", "question", "imperative", and "other". Utterance type categorization followed the convention used in the TalkBank manual. The utterance types are similar to sentence types (declarative, interrogative, imperative) with one exception: the category "question" consists of interrogatives as well as rising declaratives (i.e. declaratives with rising question intonation). In the transcripts, declaratives are marked with a period, questions with a question mark, and imperatives with an exclamation mark. It is important to note that the manual also provides terminators for special-type utterances. Among the special type utterances, this study included the following in the category "questions": trailing off of a question, question with exclamation, interruption of a question, and self-interrupted question. The category imperatives also included "emphatic imperatives". The rest of the special type utterances such as "interruptions" and "trailing off" were included in the category "other".

Results

Overall, *and* was about 10 times more likely to occur in parents' speech than *or*. More specifically, *and* occurred 15 times and *or* only 1.5 times per 1000 words. Children produced *and* at the same rate as their parents but produced *or* at a considerably lower rate, only 0.5 per thousand (**??**). The developmental trend (**??**) showed that between 12 to 72 months, production of *and* in parents' speech varied between 10 to 20 *and*'s per thousand words. Children started producing *and* between 12 and 18 months, and showed a sharp increase in their production until they reached the parent level between 30 to 36 months of age. Their productions stayed close to the parents' production level between 36 and 72 months, possibly surpassing them at 60 months – although due to the small amount of data after 60 months we should be cautious with our interpretations of the trend there. Parents produced between 1 to 2 *or*'s every thousand words. Children started producing *or* between 18 to 30 months, steadily increasing their productions until they got close to 1 *or* per thousand words at 48 months (4 years). Their *or* productions plateaued and stayed at this rate until 72 months (6 years).

Children's productions of *or* showed two main differences. First, children started producing *or* around 6 months later than *and*. Second, while children's *and* productions showed a steep rise over a year and reached the parent level around 30 months, their *or* productions rose slowly and did not reach the parent level even at 6 years of age. What factors cause these differences in children's productions of *and* and *or*? We consider three possibilities here: frequency, conceptual complexity, and usage.

First, *and* is a far more frequent connective than *or*. Goodman, Dale, and Li (2008) argue that within the same syntactic category, words with higher frequency in child-directed speech are acquired earlier. The conjunction word *and* is at least 10 times more likely to occur than *or* so earlier acquisition of *and* is consistent with the effect of frequency on age of

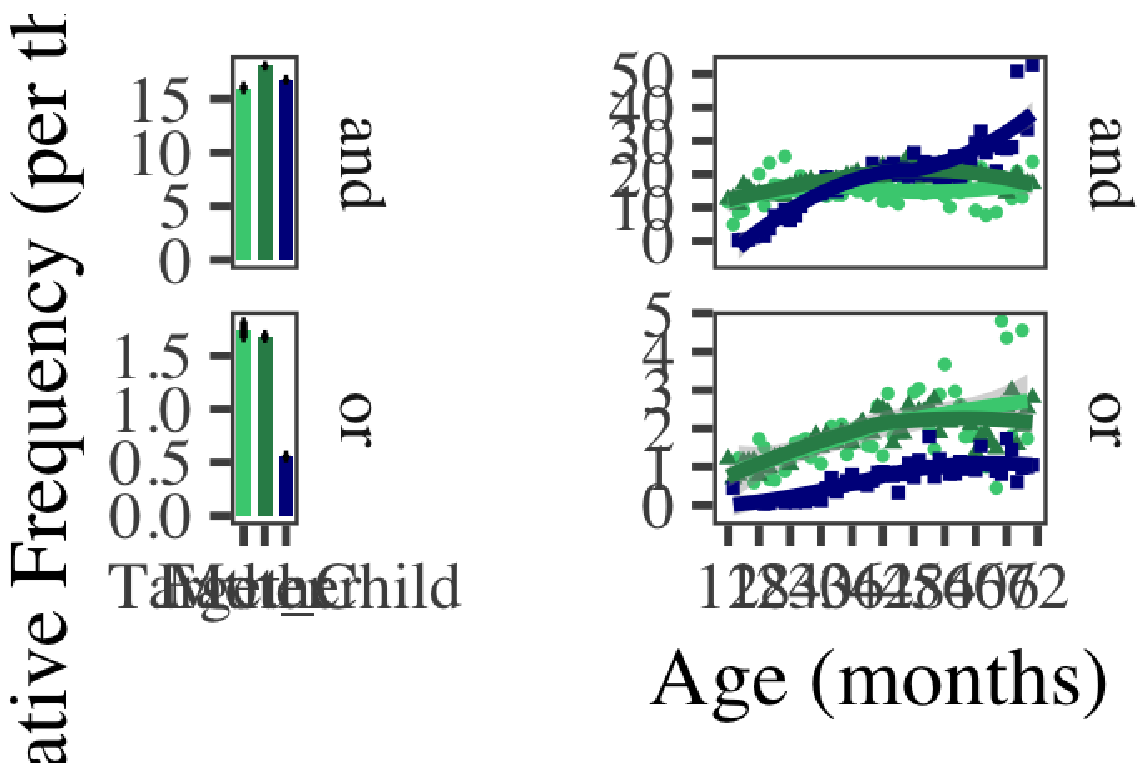


Figure 3. The relative frequency of *and/or* in the speech of fathers, mothers, and children. The y-axes show different ranges of values for *and* vs. *or*. 95% binomial proportion confidence intervals calculated using Agresti-Coull's approximate method. The monthly relative frequency of *and/or* in parents and children's speech between 12 and 72 months (1-6 years).

acquisition. Second, research on concept attainment has suggested that the concept of conjunction is easier to conjure and possibly acquire than the concept of disjunction. In experiments that participants are asked to detect the pattern of classification in some cards, they can detect a conjunctive classification faster than a disjunctive one (Neisser & Weene, 1962). Therefore, it is possible that children discover the concept that corresponds to the meaning of *and* faster and start to produce it earlier, but they need more time to attain the concept corresponding to the meaning of *or*.

A third possibility is that the developmental difference between *and* and *or* is at least partly due to their different usages. Parent-child interactions are not symmetrical and what parents would like to communicate to children is different from what children would like to

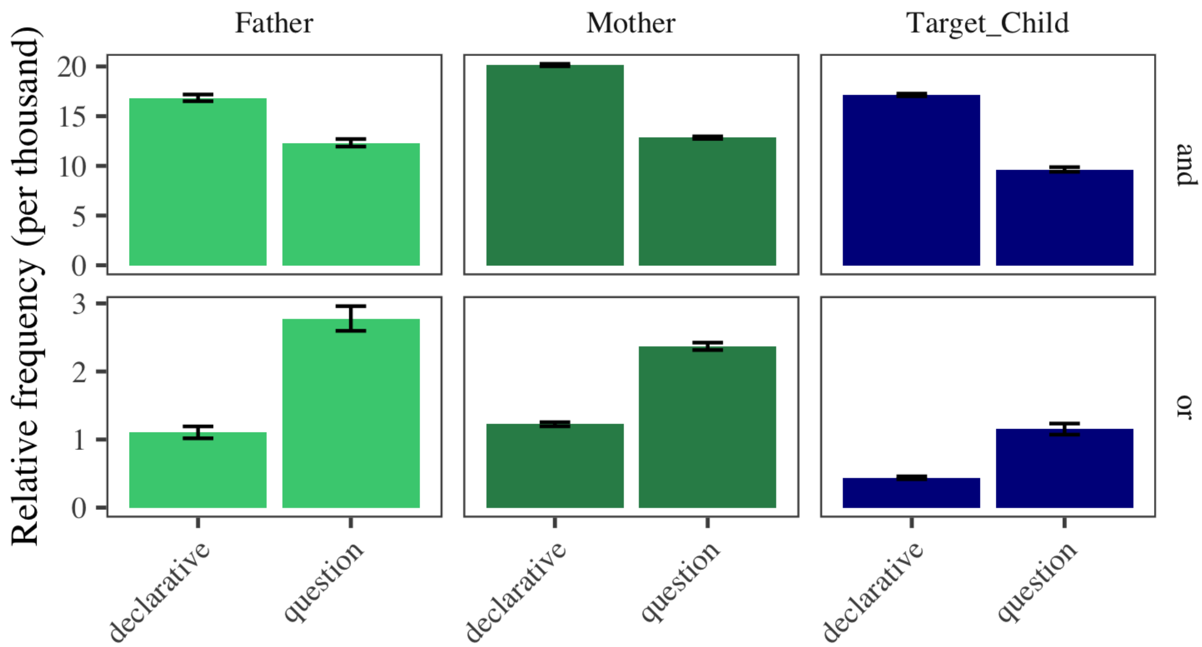


Figure 4. Relative frequency of *and/or* in declaratives, imperatives, and interrogatives for parents and children

communicate to parents. This asymmetry can result in different distribution of speech acts between parents and children and consequently functional elements that constitute them. For example, a parent saying "go to your room" to a child is not unexpected, but a child saying the same to a parent is very unexpected. Due to the nature of parent-child interactions, it is more ususal to hear parents use imperatives towards children and not vice versa. If a functional element is mostly used in imperatives, then the rate of production will be higher in parents than in children. Here we present evidence that suggests *or* is affected in a similar way.

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hla
## % Date and time: Fri, Aug 09, 2019 - 15:44:18
## \begin{table}[!htbp] \centering
```

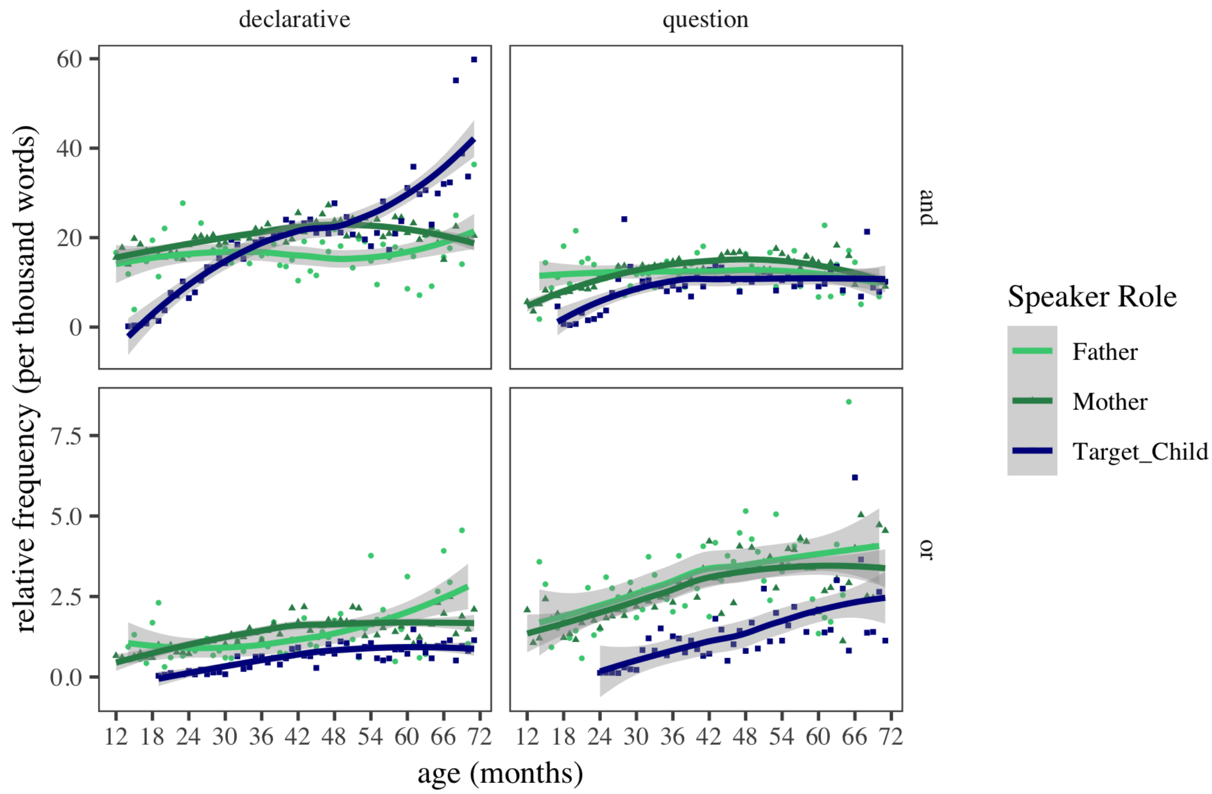


Figure 5. Relative frequency of *and/or* in declaratives and questions for parents and childern between the child-age of 12 and 72 months (1-6 years).

```
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}} \\
## \cline{2-2}
## \\[-1.8ex] & ppt \\
## \hline \\[-1.8ex]
##  target\_child\_age\_months & 0.020$^{***}$ \\
##   & (0.004) \\
```

```
##   & \\
##  speech\_actquestion & 0.601$^{**}$ \\
##   & (0.281) \\
##   & \\
##  speaker\_roleTarget\_Child & $-$0.742$^{**}$ \\
##   & (0.328) \\
##   & \\
##  target\_child\_age\_months:speech\_actquestion & 0.019$^{***}$ \\
##   & (0.006) \\
##   & \\
##  target\_child\_age\_months:speaker\_roleTarget\_Child & $-$0.001 \\
##   & (0.007) \\
##   & \\
##  speech\_actquestion:speaker\_roleTarget\_Child & $-$1.373$^{***}$ \\
##   & (0.491) \\
##   & \\
##  target\_child\_age\_months:speech\_actquestion:speaker\_roleTarget\_Child & 0.013 \\
##   & (0.010) \\
##   & \\
##  Constant & 0.521$^{***}$ \\
##   & (0.199) \\
##   & \\
## \hline \\[-1.8ex]
## Observations & 220 \\
## R$^{2}$ & 0.721 \\
## Adjusted R$^{2}$ & 0.712 \\
## Residual Std. Error & 0.593 (df = 212) \\
```

```
## F Statistic & 78.342$^{***}$ (df = 7; 212) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01
## \end{tabular}
## \end{table}
```

First, we found that *or* is more likely to occur in questions than in declaratives (4). This is in contrast to *and* which is more likely to occur in declaratives. Second, parents asked more questions from children than children did from parents, and children produced more declaratives than parents (6). In fact, questions have their own developmental trajectory, emerging in the second year of children's lives and reaching a relatively constant rate of about 15% of children's utterances in their fourth year. However, parents produce a constant rate of questions which is about 25% of their utterances. Therefore, parent-child interaction provides more opportunities for parents to use *or* than children.

Figure 5 shows the developmental trends of the relative frequencies of *and* and *or* in questions and declaratives. Comparing *and* in declaratives and questions, we see that the onset of *and* productions are slightly delayed for questions but in both declaratives and questions, *and* productions reach the parent level around 36 months (3 years). For *or*, we see a similar delay in questions compared to declaratives. Children start producing *or* in declaratives at around 18 months but they start producing *or* in questions at 24 months. Production of *or* increases in both declaratives and questions until it seems to reach a constant rate in declaratives between 48 and 72 months. The relative frequency of *or* in questions continues to rise until 60 months. Comparing figures **??** and 5, we see that children are closer to the adult rate of production in declaratives than questions.

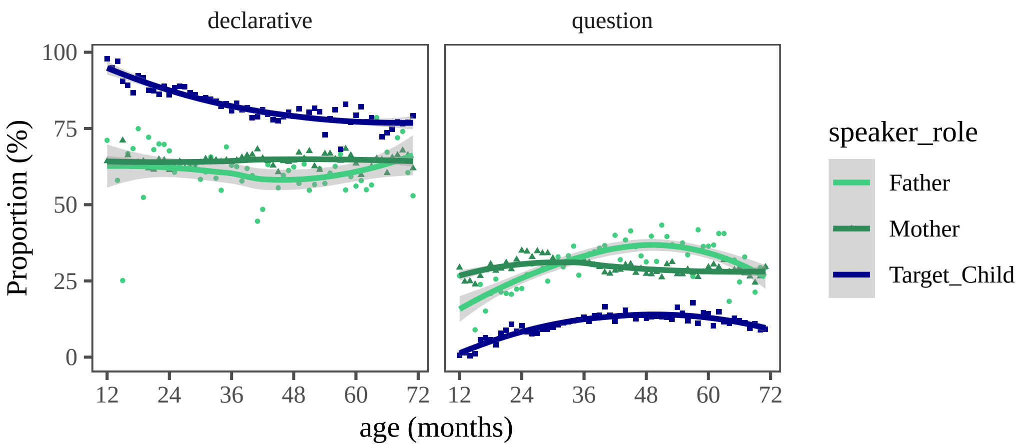


Figure 6. Proportion of declaratives to questions in parent-child interactions by age.

Conclusion

In a large-scale quantitative analysis of parents and children's productions of *and* and *or*, we found that children started producing *and* in the second year of their lives, and quickly reached their parents' rate of production by two and a half. Their production of disjunction was delayed by six months on average: they started producing *or* between 1.5 and 2.5 years of age, and around 3.5 years, they reached a relatively constant rate of production below that of their parents. We considered three possible causes for disjunction's delay and lower rate of production: the higher frequency of *and*, the conceptual and mapping complexity of *or*, and the asymmetry in speech acts produced by parents and children. We provided evidence for the last cause. We showed that parents produced more questions than children, and that *or* was more likelly to occur in questions. Therefore, parents' speech contained more *or* partly due to the fact that parents asked more questions.

Study 2: Interpretations of disjunction in child-directed speech

In this study we selected a subset of connective examples in child-directed speech from study 1 to closely examine the interpretations they recieve. Research in formal semantics has shown that the interpretation of disjunction depends on several including prosody (Pruitt & Roelofsen, 2013), logical consistency of the propositions being connected (Geurts, 2006). Our main claim here is that in child-directed speech, exclusive interpretations of “or” correlate with rise-fall prosody and logically inconsistent propositions. In the absence of these two factors, *or* is most likely not exclusive.

Methods

This study used the Providence corpus (Demuth, Culbertson, & Alter, 2006) available via the PhonBank section of the TalkBank.org archive. The corpus was chosen because of its relatively dense data on child-directed speech as well as the availability of audio and video recordings that would allow annotators access to the context of the utterance. The corpus was collected between 2002 and 2005 in Providence, Rhode Island. Table 2 in appendix reports the name, age range, and the number of recording sessions for the children in this study. All children were monolingual English speakers and were followed between the ages of 1 and 4 years. Based on Study 2, this is the age range when children develop their early understanding of *and* and *or*. The corpus contains 364 hours of biweekly hour-long interactions between parents and children.

Exclusion Criteria. We excluded data from Ethan since he was diagnosed with Asperger’s Syndrome at age 5. We also excluded all examples found in conversations over the phone, adult-adult conversations, and utterances heard from TV or radio. We did not count such utterances as child-directed speech. We excluded proper names and fixed forms such as “Bread and Circus” (name of a local place) or “trick-or-treat” from the set of examples to be annotated. Such forms could be learned and understood with no actual understanding of the connective meaning. We counted multiple instances of *or* and *and* within the same disjunction/conjunction as one instance. The reasoning was that, in a coordinated structure, the additional occurrences of a connective typically did not alter the annotation categories, and most importantly the interpretation of the coordination. For example, there is almost no difference between "cat, dog, and elephant" versus "cat and dog and elephant" in interpretation. In short, we focused on the "coordinated construction" as a unit rather than on every separate instance of *and* and *or*. Instances of multiple connectives in a coordination were rare in the corpus.

Procedure. All utterances containing *and* and *or* were extracted using the CLAN software and automatically tagged for the following: (1) the name of the child; (2) the transcript address; (3) the speaker of the utterance (father, mother, or child); (4) the child's birth date, and (5) the recording date. Since the focus of the study was mainly on disjunction, we annotated instances of *or* in all the child-directed speech from the earliest examples to the latest ones found. Given that the corpus contained more than 10 times the number of *and*'s than *or*'s, we randomly sampled 1000 examples of *and* to match 1000 examples of *or*. Here we report the results on 465 examples of *and* and 608 examples of *or*.

Annotation Categories. Every extracted instance of *and* and *or* was manually annotated for 7 categories: connective interpretation, intonation type, utterance type, syntactic level, conceptual consistency, communicative function, and answer type. We briefly explain how each annotation category was defined. Further details and examples are provided in the appendix section.

1. *Connective Interpretation*

This annotation category was the dependent variable of the study. Annotators listened to coordinations such as "A or B" and "A and B", and decided the intended interpretation of the connective with respect to the truth of A and B. We used the sixteen binary connectives shown in Figure 36 as the space of possible connective interpretations. Annotators were asked to consider the two propositions raised by the coordinated construction, ignoring the connective and functional elements such as negation and modals. Consider the following sentences containing *or*: "Bob plays soccer or tennis" and "Bob doesn't play soccer or tennis". Both discuss the same two propositions: A. Bob playing soccer, and B. Bob playing tennis. However, the functional elements combining these two propositions result in different interpretations with respect to the truth of A and B. In "Bob plays soccer or tennis" which contains a disjunction, the interpretation is that Bob plays one or possibly both sports (IOR). In "Bob doesn't play soccer or tennis" which contains a negation and a disjunction, the interpretation is that Bob plays neither sport (NOR). For connective interpretations, the annotators first reconstructed the coordinated propositions without the connectives or negation and then decided which propositions were implied to be true/false.

2. *Intonation Type*

Annotators listened to the utterances and decided whether the intonation contour on the coordination was flat, rise, or rise-fall. Table 4 in the appendix shows the definitions and examples for these intonation types. In order to judge the intonation of the sentence accurately, annotators were asked to construct all three intonation contours for the same sentence and see which one is closer to the actual intonation of the utterance. For example, to judge the sentence "do you want orange juice↑ or apple juice↓?", they reconstructed the sentence with the prototypical flat, rising, and rise-fall intonations and checked to see which intonation is closer to the actual one.

3. *Utterance Type*

Annotators decided whether an utterance was an instance of a declarative, an interrogative, or an imperative. Occasionally, we found examples with different utterance types for each coordinand. For example, a mother could say "put your backpack on and I'll be right back", where the first cooridnand is an imperative and the second a declarative. Such examples were coded for both utterance types with a dash inbetween: imperative-declarative. Table 5 in the appendix provides the detailed definitions and examples for each utterance type.

4. *Syntactic Level*

Annotators marked whether the coordination was at the clausal level or at the sub-clausal level. Clausal level was defined as sentences, clauses, verb phrases, and verbs. Coordination of other categories was coded as sub-clausal. This annotation category was introduced to check the hypothesis that the syntactic category of the coordinands may influence the interpretation of a coordination. For example, a sentence like "He drank tea or coffee" is less likely to be interpreted as exclusive than "He drank tea or he drank coffee." The clausal vs. sub-clausal distinction was inspired by the fact that in many languages, coordinators that connect sentences and verb phrases are different lexical items than those that connect nominal, adjectival, or prepositional phrases (see Haspelmath, 2007).

4. *Conceptual Consistency*

Propositions stand in complex conceptual relations with each other. For example, have logical, temporal, and causal relation with each other. For conceptual consistency, annotators decided whether the propositions that made up the coordination could be true at the same time or not. If the two propositions could not be true at the same time and resulted in a contradiction, they were marked as inconsistent. Our annotators used the following diagnostic to decide the consistency of the disjuncts: Two disjuncts were marked as inconsistent if replacing the word *or* with *and* produced a contradiction. For example, changing "the ball is in my room *or* your room" to "the ball is in my room *and* your room" produces a contradiction because a ball cannot be in two rooms at the same time.

It is important to discuss two issues regarding conceptual consistency. First, our diagnostic for consistency was quite strict. In many cases, propositions are not inconsistent in this sense but they are implausible. For example, drinking both tea and coffee at the same time is not inconsistent, but is unlikely. It is possible that many exclusive interpretations are based on such judgments of implausability. Second, if the coordinands are inconsistent, this does not necessarily mean that the connective interpretation must be exclusive. For example, in a sentence like “you could stay here or go out”, the alternatives “staying here” and “going out” are inconsistent. Yet, the overall interpretation of the connective could be conjunctive: you could stay here AND you could go out. The statement communicates that both possibilities hold. This pattern of interaction between possibility modals like *can* and disjunction words like *or* are often discussed under “free-choice inferences” in the semantics and pragmatics literature (Kamp, 1973; Von Wright, 1968). Another example is unconditionals such as “Ready or not, here I come!”. The coordinands are contradictions: one is the negation of the other. However, the overall interpretation of the sentences is that in both cases, the speaker is going to come.

5. *Communicative Functions*

We constructed a set of categories that captured particular usages or communicative functions of the words *or* and *and*. They include descriptions, directives, preferences, identifications, definitions-examples, clarifications, repairs, and a few others shown in Table 8 in appendix. These communicative functions were created using the first 100 examples and then they were used for the classification of the rest of the examples. Some communicative functions are general and some are specific to coordination. For example, directives are a general class while conditionals (e.g. Put that out of your mouth, or I’m gonna put it away) are more specific to coordinated constructions. It is also important to note that the list is not unstructured. Some communicative functions are subtypes of others. For example, “identifications” and “unconditionals” are subtypes of “descriptions” while “conditionals” are a subtype of directives. Furthermore, “repairs” seem parallel to other categories in that any type of speech can be repaired. We do not fully explore the details of these functions in this study but such details matter for a general theory of acquisition that makes use of the speaker’s communicative intentions as early coarse-grained communicative cues for the acquisition of fine-grained meaning such as function words.

6. *Answer Type*

Whenever a parent's utterance was a polar question, the annotators coded the utterance for the type of response it received from the children. This annotation category was different from others in that it was included as a cue for learning disjunction. It was used as an opportunity to assess, albeit in a limited and indirect way, the comprehension of children in the same corpus. Table 9 in the appendix shows the answer types in this study and their definitions and examples. Utterances that were not polar questions were simply coded as NA for this category. If children responded to polar questions with "yes" or "no", the category was YN and if they repeated with one of the coordinands the category was AB. If children said yes/no and followed it with one of the coordinands, the answer type was determined as YN (yes/no). For example, if a child was asked "Do you want orange juice or apple juice?" and the child responded with "yes, apple juice", our annotators coded the response as YN. The reason is that in almost all cases, if a simple yes/no response is felicitous, then it can also be optionally followed with mentioning a disjunct. However, if yes/no is not a felicitous response, then mentioning one of the alternatives is the only appropriate answer. For example, if someone asks "Do you want to stay here or go out?" a response such as "yes, go out" is infelicitous and a better response is simply "go out". Therefore, we counted responses with both yes/no and mentioning an alternative as a yes/no response.

Inter-annotator Reliability. To train annotators and confirm their reliability for disjunction examples, two annotators coded the same 240 instances of disjunction. The inter-annotator reliability was calculated over 8 iterations of 30 examples each. After each iteration, annotators met to discuss disagreements and resolve them. They also decided whether the category definitions or annotation criteria needed to be made more precise. Training was completed after three consecutive iterations showed substantial agreement between the annotators for all categories (Cohen's $\kappa > 0.7$). Further details on inter-annotator reliability are presented in the appendix section.

Results

Figure blah shows the distribution of connective interpretations in child-directed speech. The most common interpretation was the conjunctive interpretation (AND, 49%) followed by the exclusive interpretation (XOR, 35%). The figure also shows the distribution of connective interpretations broken down by the connective word used: *and* vs. *or*[1]. The most frequent interpretation for *and* was conjunction (AND), and for *or*, exclusive disjunction (XOR). These results replicated the findings of Morris (2008).

Figure 9 shows connective interpretations by the syntactic level of the disjunction. The results suggest a small effect of clausal level disjuncts. Disjunctions were more likely to be interpreted as exclusive when their disjuncts were clauses or verbs rather than nominals, adjectives, or prepositions (all sub-clausal units).

Finally, figure 10 shows the proportions of connective interpretations in the 10 different communicative functions we defined. The results show that certain functions increase the likelihood of some connective interpretations. An exclusive (XOR) interpretation of *or* is common in acts of clarification, identification, stating/asking preferences, stating/asking about a description, or making a conditional statements. These results are consistent with expectations on the communicative intentions of that these utterances carry. In clarifications, the speaker needs to know which of two alternatives the other party meant. Similarly in identifications, speaker needs to know which category does a referent belongs to. In preferences, parents seek to know which of two alternatives the child wants. Even though descriptions could be either inclusive or exclusive, in the current sample, most descriptions were questions about the state of affairs and required the child to provide one of the

[1] All the confidence intervals shown in the plots for this section are simultaneous multinomial confidence intervals computed using the Sison and Glaz (1995) method.

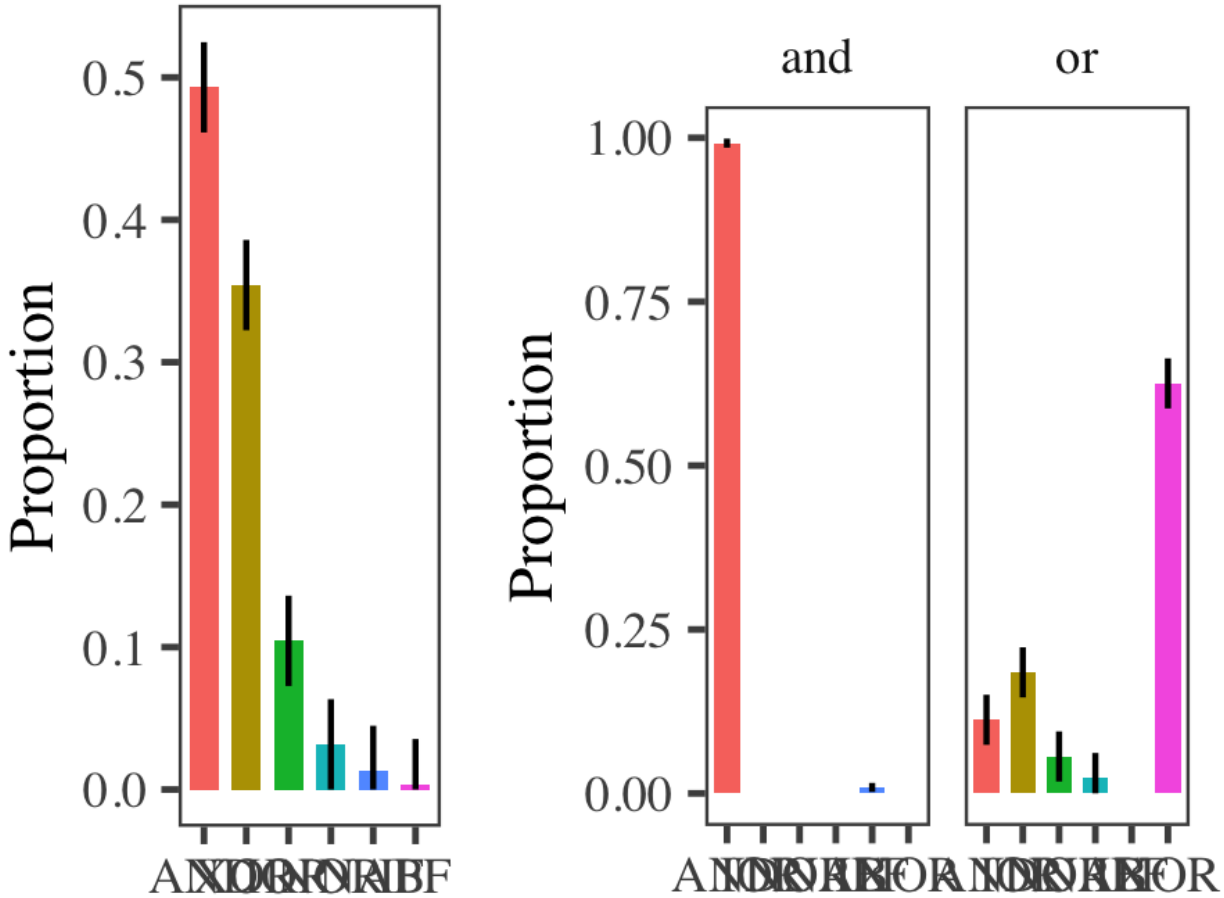


Figure 7. The proportion of different interpretations of the connectives *and/or* in child-directed speech. Interpretations of *and/or* in child-directed speech

alternatives as the answer. In conditionals such as “come here or you are grounded”, the point of the threat is that only one disjunct can be true: either “you come and you are not grounded” or “you don’t come and you are grounded”.

Repairs often received an exclusive (XOR) or a second-disjunct-true (NAB) interpretation. This is expected given that in repairs the speaker intends to say that the first disjunct is incorrect or inaccurate. Unconditionals and definitions/examples always had a conjunctive (AND) interpretation. Again, this is to be expected. In such cases the speaker

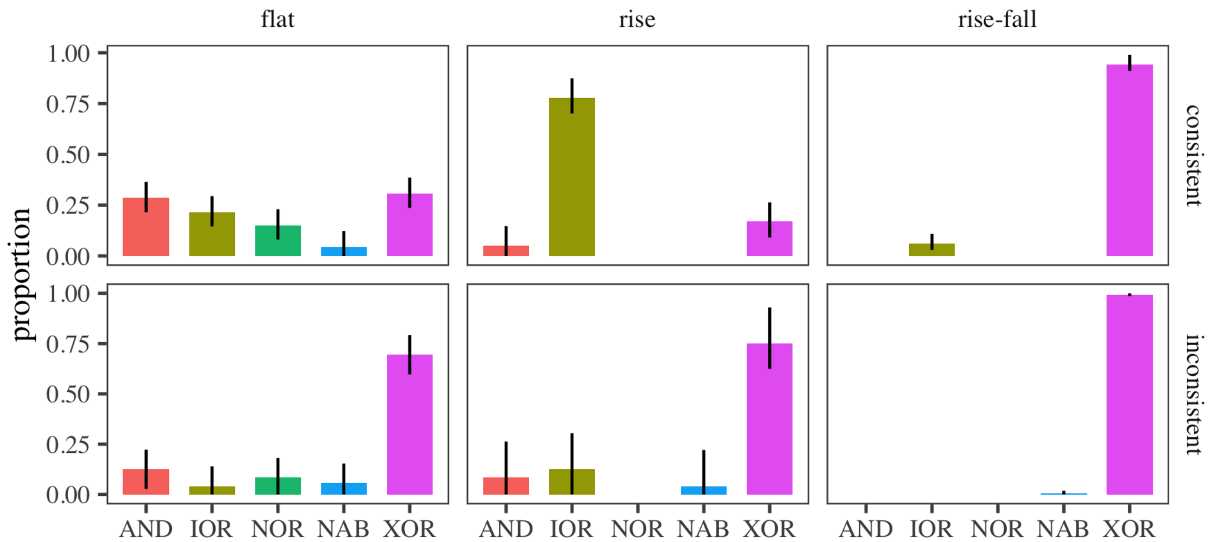


Figure 8. Interpretations of and/or in the three intonation contours flat, rising, and rise-fall.

intends to communicate that all options apply. If the mother says that “cats are animals like lions or tigers”, she intends to say that both lions and tigers are cats, and not one or the other. Interestingly, in some cases (not all), *or* is replaceable by *and*: “cats are animals like lions and tigers”. In unconditionals, the speaker communicates that in both alternatives, a certain proposition holds. For example, if the mother says “ready or not, here I come!”, she communicates that “I come” is true in both cases where “you are ready” and “you are not ready”.

Options were often interpreted either as conjunctive (AND) or inclusive (IOR). The category “options” contained examples of free-choice inferences such as “you could drink orange juice or apple juice”. This study found free-choice examples much more common than the current literature on the acquisition of disjunction suggests. Finally, directives received either an IOR or XOR interpretation. It is important to note here that the most common communicative function in the data were preferences and descriptions. Other communicative functions such as unconditionals or options were fairly rare. Despite their infrequent appearance, these constructions must be learned by children at some point, since almost all

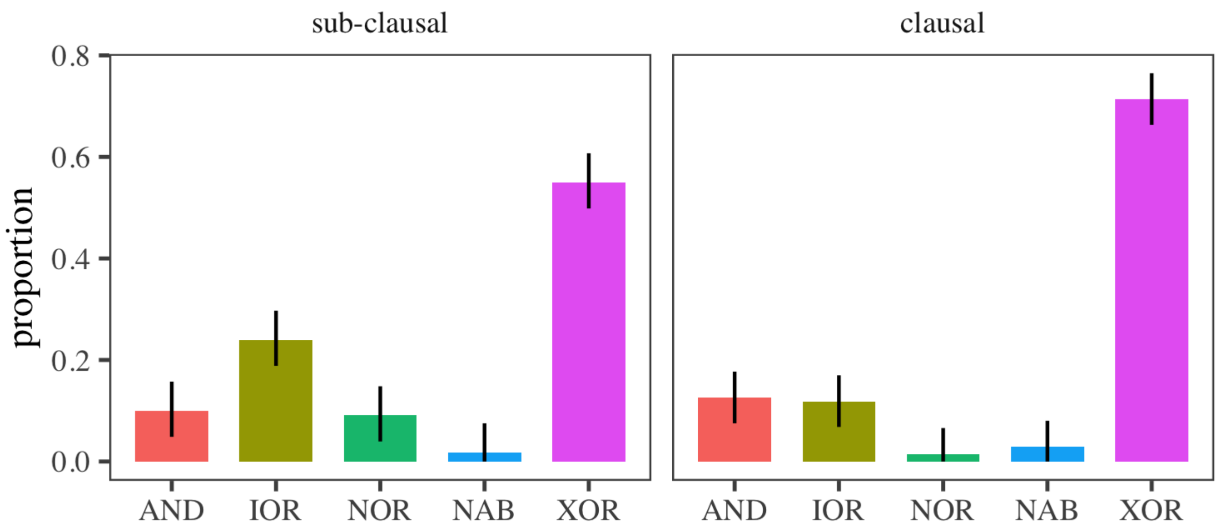


Figure 9. Connective interpretations in clausal and sub-clausal disjunctions.

adults know how to interpret them. It is clear from the investigation here that any learning account for function word meaning/interpretation also needs to account for how such infrequent constructions are learned.

Comprehension. First we look at how children responded to their parents' questions with *or* (Answer Type). Figure 11 shows the monthly proportions of "yes/no" and alternative (AB) answers between the ages of 1 and 3 years. Initially, children provided no answer to questions, but by the age of 3 years, the majority of such questions received a yes/no (YN) or alternative (AB) answer. This increase in the proportion of responses to questions containing *or* between 20 to 30 months of age suggests that initial form-meaning mappings for disjunction is formed in this age range.

Conclusion

This study focused on the interpretations that connectives *and* and *or* recieve in child-directed speech. It also investigated some candidate cues that can help children learn

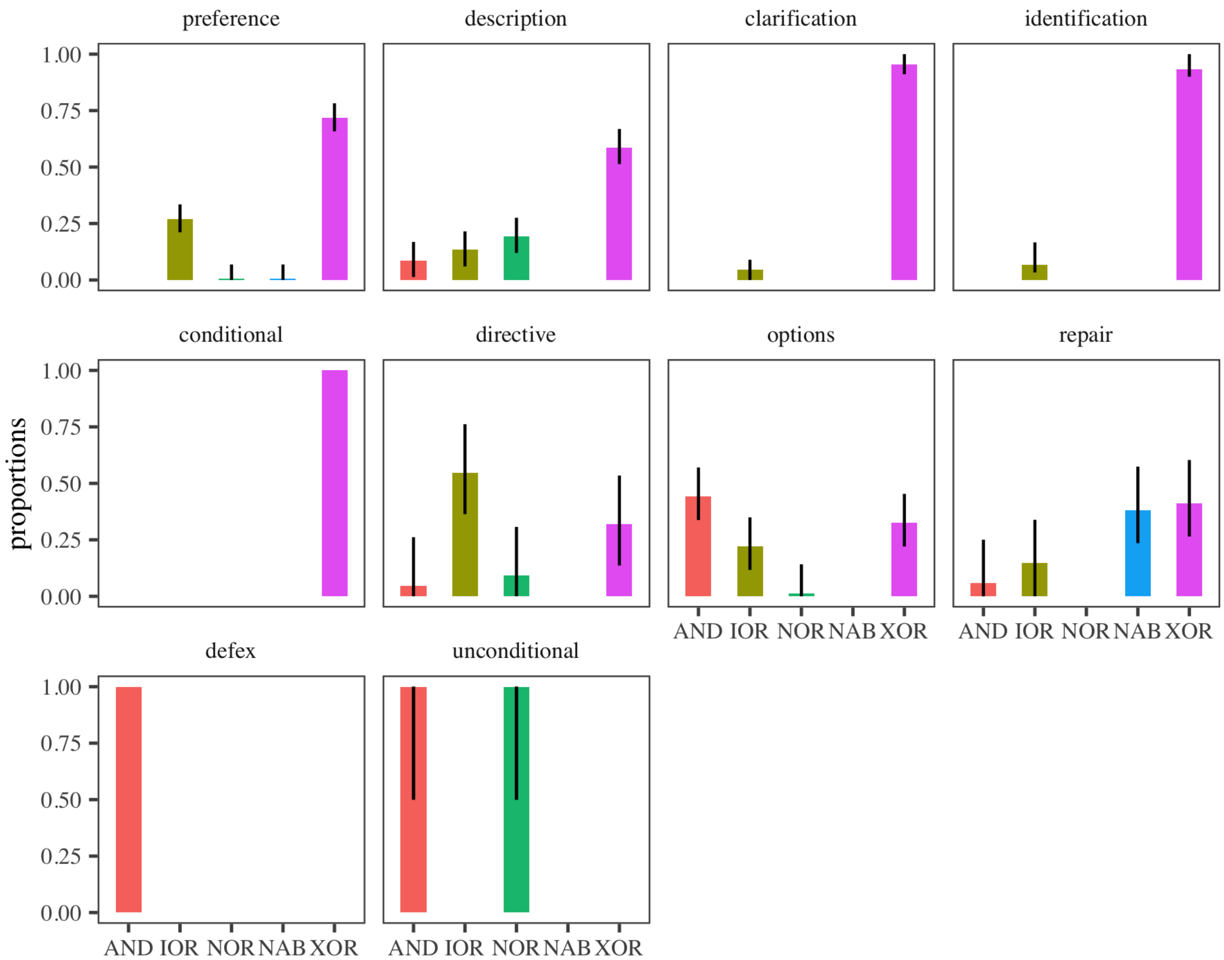


Figure 10. Connective interpretations in different communicative functions.

these interpretations of a disjunction. The study presented 1000 examples of *and* and *or* in child-directed speech, annotated for their truth-conditional interpretation, as well as five candidate cues to their interpretation: (1) Utterance Type; (2) Intonation Type; (3) Syntactic Level; (4) Conceptual Consistency; and (5) Communicative Function. Like Morris (2008), this study found that the most common interpretations of *and* and *or* are conjunction (AND) and exclusive disjunction (XOR) respectively. However, we found many inclusive and conjunctive instances of *or* as well.

The most likely interpretation of a disjunction depended on the cues that accompanied

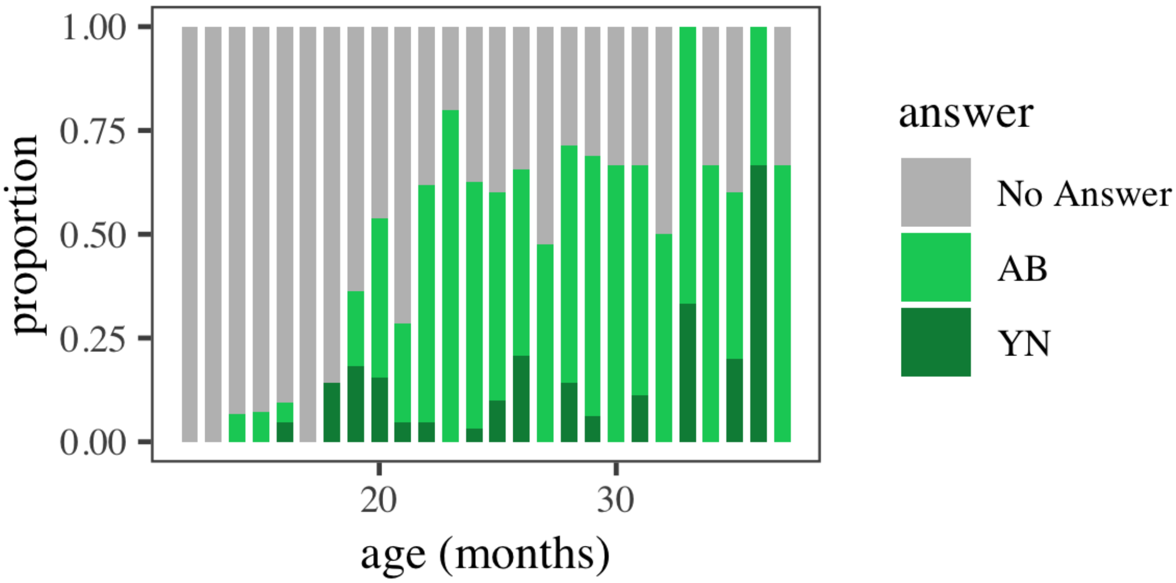


Figure 11. The proportions of children's answer types to polar questions containing the connective *or* at different ages (in months).

it. A disjunction was most likely exclusive if the alternatives were inconsistent (i.e. contradictory). A disjunction was either inclusive or exclusive if it appeared in a question. Within questions, a disjunction was most likely exclusive if the intonation was rise-fall. If the intonation was rising, the question was interpreted as inclusive. The syntactic category of the disjuncts could also provide information for interpretation. If the disjuncts were clausal then it was more likely for the disjunction to be interpreted as exclusive, even though this effect was small. Finally, specific communicative functions required specific interpretations of the connective. *Or* often received a conjunctive interpretation in the following contexts: defining terms and providing examples, enumerating options, and in unconditional constructions. These results suggest that a learner can rely on cues that accompany a disjunciton for its interpretation. In the next section, we develop a computational model to test this hypothesis more formally.

Study 3: Learning to interpret a disjunction

Given the wide range of interpretations that *or* can have, how can children learn to interpret it correctly? This is what study @ref addresses. In doing so, it also provides a solution to the puzzle of learning disjunction. To remind you about the puzzle, previous research have shown that the majority of *or*-examples children hear are exclusive. However, comprehension studies report that between the ages of three and five, children can interpret *or* as inclusive disjunction in declarative sentences (Crain, 2012). The finding of the comprehension studies and the corpus studies taken together present a learning puzzle: how can children learn to interpret *or* as inclusive if they mostly hear exclusive examples? This study provides a solution by developing a cue-based account for children's acquisition of connectives. More generally, the account proposed is helpful for learning words with multiple interpretations when one interpretation dominates the learner's input.

Cues to coordinator meanings

Three important compositional cues can help learners in restricting their hypotheses to coordinator meanings. First, as pointed out by Haspelmath (2007), coordination has specific compositional properties. Coordinators combine two or more units of the same type and return a larger unit of the same type. The larger unit has the same semantic relation with the surrounding words as the smaller units would have had without coordination. These properties separate coordinators from other function words such as articles, quantifiers, numerals, prepositions, and auxiliaries which are not used to connect sentences or any two similar units for that matter. In fact, the special syntactic properties of coordinators have compelled syntactic theories to consider specific rules for coordination.

The literature on syntactic bootstrapping suggests that children can use syntactic properties of the input to limit their word meaning hypotheses to the relevant domain

(Brown, 1957; Gleitman, 1990; see Fisher, Gertner, Scott, & Yuan, 2010 for a review). In the current 1073 annotations of conjunction and disjunction, we found that *and* and *or* connected sentences/clauses 56% of the time. This pattern is unexpected for any other class of function words and it is possible that the syntactic distribution of coordinators cue the learners to the space of sentential connective meanings.

Second, in the annotation study we found that *and* never occurs with inconsistent coordinands (e.g. "clean and dirty") while *or* commonly does (e.g. "clean or dirty"). The inconsistency of the coordinands can cue the learner to not consider conjunction as a meaning for the coordinator given that a conjunctive meaning would too often lead to a contradiction at the utterance level. On the other hand, choosing disjunction as the meaning avoids this problem. Third, Study 1 found that *or* is more likely to occur in questions than statements while *and* is more likely in statements. Since questions often contain more uncertainty while statements are more informative, it is possible that these environments bias the learner towards selecting hypotheses that match this general communicative function. Disjunction is less informative than conjunction and it is possible that the frequent appearance of *or* in questions cues learners to both its meaning as a disjunction as well as the ignorance inference commonly associated with it.

Finally, it is reasonable to assume that not all binary connective meanings shown in Figure 12 are as likely for mapping. For example, coordinators that communicate tautologies or contradictions seem to be not good candidates for informative communication. Similarly, if A coordinated with B simply asserts the truth of A and says nothing about B, it is unclear why it would be needed if the language already has the means of simply asserting A. It is possible that pragmatic principles already bias the hypothesis space to favor candidates that are communicatively more efficient.

Even though these findings are suggestive, they need to be backed up by further observational and experimental evidence to show that children do actually use these cues in

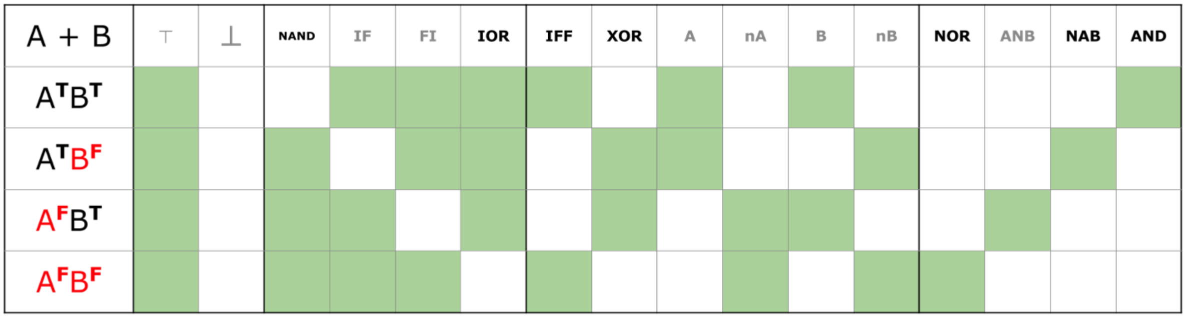


Figure 12. The truth table for the 16 binary logical connectives. The rows represent the set of situations where zero, one, or both propositions are true. The columns represent the 16 possible connectives and their truth conditions. Green cells represent true situations.

learning connective meanings. In the next section, I turn to the more specific issue of learning the correct interpretation of *and* and *or* from the input data. As in the case of number words, previous research has provided insight into how children comprehend a disjunction and what they hear from their parents. The main question is how children learn what they comprehend from what they hear. I turn to this issue in the next section.

Learning to interpret *and* and *or*: A cue-based account

Previous comprehension studies have shown that children as early as age three can interpret a disjunction as inclusive (see Crain, 2012 for an overview). However, Morris (2008) showed that exclusive interpretations are much more common than other interpretations of disjunction in children's input. Figure 13 shows the results of our annotation study by grouping the disjunction interpretations into exclusive (EX) and inclusive (IN), i.e. non-exclusive categories. These results replicate Morris' (2008) finding and reinforce a puzzle raised by Crain (2012): How can children learn the inclusive interpretation of disjunction when the majority of the examples they hear are exclusive? To answer this question, we draw on insights from the Gricean approach to semantics and pragmatics.

Research in Gricean semantics and pragmatics has shown that the word *or* is not the only factor relevant to the interpretation of a disjunction. It is not only the presence of the word *or* that leads us to interpret a disjunction as inclusive, exclusive, or conjunctive, but rather the presence of *or* along with several other factors such as intonation (Pruitt & Roelofsen, 2013), the meaning of the disjuncts (Geurts, 2006), and the conversational principles governing communication (Grice, 1989). The interpretation and acquisition of the word *or* cannot, therefore, be separated from all the factors that accompany it and shape its final interpretation.

In the literature on word learning and semantic acquisition, form-meaning mapping is often construed as mapping an isolated form such as *gavagai* to an isolated concept such as "rabbit". While this approach may be feasible for content words, it will not work for function words such as *or*. First, the word *or* cannot be mapped in isolation from its formal context. As Pruitt and Roelofsen (2013) showed, the intonation that accompanies a disjunction affects its interpretation. Therefore, a learner needs to pay attention to the word *or* as well as the intonation contour that accompanies it. Second, the word *or* cannot be mapped to its meaning isolated from the semantics of the disjuncts that accompany it. As Geurts (2006) argued, the exclusive interpretation is often enforced simply because the options are incompatible. For example, "to be or not to be" is exclusive simply because one cannot both be and not be. In addition, conversational factors play an important role in the interpretation of *or* as Grice (1989) argued. In sum, the interpretation and acquisition of function words such as *or* require the learner to consider the linguistic and nonlinguistic context of the word and map the meanings accordingly.

Previous accounts have adopted a model in which a function word such as *or* is mapped directly to its most likely interpretation:

$or \rightarrow \oplus$

This model is often used in cross-situational accounts of content words. Here I argue that the direct mapping of *or* to its interpretation without consideration of its linguistic context is the primary cause of the learning puzzle for *or*. Instead, I propose that the word *or* is mapped to an interpretation in a context-dependent manner, along with the interpretive cues that accompany it such as intonation and disjunct semantics:

[connective: *or*, Intonation: rise-fall, Disjuncts: inconsistent] $\rightarrow \oplus$

[connective: *or*, Intonation: rising, Disjuncts: consistent] $\rightarrow \vee$

Figure 14 shows that the rate of exclusive interpretations change systematically when the data are broken down by intonation and consistency. Given a rise-fall intonation contour, a disjunction is almost always interpreted as exclusive. Similarly, if the propositions are inconsistent, the disjunction is most likely interpreted as exclusive. When either of these two features are absent, a disjunction is more likely to receive an inclusive interpretation.

In this account, it is not a single word that gets mapped to an interpretation but rather a cluster of features. This method has two advantages. First, it deals with the context dependency of disjunction interpretation. The learner knows that *or* with some intonation has to be interpreted differently from one with another. Second, it allows the learner to pull apart the contribution of *or* from the interpretive cues that often accompany it. In fact, analysis of all mapping clusters in which *or* participates and generalization over them can help the learner extract the semantics of *or* the way it is intended by Gricean accounts of semantics/pragmatics. For those skeptical of such an underlying semantics for *or*, there is no need for further analysis of the mapping clusters. The meaning of *or* as a single lexical item is distributed among the many mappings in which it participates. In the next section, I implement this idea using decision tree learning.

A decision tree is a classification model structured as a hierarchical tree with nodes, branches, and leaves (Breiman, 2017). The tree starts with an initial node, called the root, and branches into more nodes until it reaches the leaves. Each node represents the test on a feature, each branch represents an outcome of the test, and each leaf represents a classification label. Using a decision tree, observations can be classified or labeled based on a set of features.

Decision trees have several advantages for modeling cue-based accounts of semantic acquisition. First, decision trees use a set of features to predict the classification of observations. This is analogous to using cues to predict the correct interpretation of a word or an utterance. Second, unlike many other machine learning techniques, decision trees result in models that are interpretable. Third, the order of decisions or features used for classification is determined based on information gain. Features that appear higher (earlier) in the tree are more informative and helpful for classification. Therefore, decision trees can help us understand which cues are probably more helpful for the acquisition and interpretation of a word.

Decision tree learning is the construction of a decision tree from labeled training data. This section applies decision tree learning to the annotated data of Study 3 by constructing random forests (Breiman, 2001; Ho, 1995). In random forest classification, multiple decision trees are constructed on subsets of the data, and each tree predicts a classification. The ultimate outcome is a majority vote of each trees classification. Since decision trees tend to overfit data, random forests control for overfitting by building more trees and averaging their results. **(Citation)** Next section discusses the methods used in constrcting the random forests for interpreting connectives *or/and*.

Methods. The random forest models were constructed using python's Sci-kit Learn package (Pedregosa et al., 2011). The annotated data had a feature array and a connective interpretation label for each connective use. Connective interpretations included exclusive

(XOR), inclusive (IOR), conjunctive (AND), negative inclusive (NOR), and NPQ which states that only the second proposition is true. The features or cues used included all other annotation categories: intonation, consistency, syntactic level, utterance type, and communicative function. All models were trained with stratified 10-Fold cross-validation to reduce overfitting. Stratified cross-validation maintains the distribution of the initial data in the random sampling to build cross validated models. Maintaining the data distribution ensures a more realistic learning environment for the forests. Tree success was measured with F1-Score, harmonic average of precision and recall **(Citation)**.

First a grid search was run on the hyperparamter space to establish the number of trees in each forest and the maximum tree depth allowable. The grid search creates a grid of all combinations of forest size and tree depth and then trains each forest from this grid on the data. The forests with the best F1-score and lowest size/depth are reported. **(Citation*) **The default number of trees for the forests was set to 20, with a max depth of eight and a minimum impurity decrease of 0. Impurity was measured with gini impurity, which states the odds that a random member of the subset would be mislabled if it were randomly labeled according to the distribution of labels in the subset.** (Citation)**

Decision trees were fit with high and low minimum gini decrease values. High minimum gini decrease results in a tree that does not use any features for branching. Such a tree represents the baseline or traditional approach to mapping that directly maps a word to its most likely interpretation. Low minimum gini decrease allows for a less conservative tree that uses multiple cues/features to predict the interpretation of a disjunction. Such a tree represents the cue-based context-sensitive account of word learning discussed in the previous section.

Results. We first present the results of the random forests in the binary classification task. The models were trained to classify exclusive and inclusive interpretations of disjunction. For visualization of trees, we selected the highest performing tree in the forest by testing each tree and selecting for highest F1 score. While the forests performance is not identical to the highest performing tree, the best tree gives an illustrative example of how the tree performs.

Figure 15 shows the best performing decision tree with high minimum gini decrease. As expected, a learner that does not use any cues would interpret *or* as exclusive all the time. This is the baseline model. Figure 16 shows the best performing decision tree with low minimum gini decrease. The tree has learned to use intonation and consistency to classify disjunctions as exclusive or inclusive. As expected, if the intonation is rise-fall or the disjuncts are inconsistent, the interpretation is exclusive. Otherwise, the disjunction is classified as inclusive.

Figure 17 shows the average F1 scores of the baseline and cue-based models in classifying exclusive examples. The models perform relatively well and similar to each other, but the cue-based model performs slightly better. The real difference between the baseline model and the cue-based model is in their performance on inclusive examples. Figure 18 shows the F1 score of the forests as a function of the training size in classifying inclusive examples. As expected, the baseline model performs very poorly while the cue-based model does a relatively good job and improves with more examples.

Next, we use decision tree learning in a ternary classification task. The model uses features to interpret a coordination with *and* and *or* as inclusive (IOR), exclusive (XOR), or conjunctive (AND). Figure 19 shows the baseline decision tree with high minimum gini decrease, which only uses the presence of the words *or*/*and* to interpret conjunction and disjunction. As expected, the tree interprets a coordination with *and* as a conjunction and one with *or* as exclusive disjunction. Figure 20 shows the cue-based decision tree with low minimum gini decrease. In addition to the presence of *and* and *or*, the tree uses intonation, consistency, communicative function, and utterance type to distinguish exclusive, inclusive, and conjunctive uses of disjunction. In short, a disjunction that is rise-fall, inconsistent, or has a conditional communicative function is classified as exclusive. Otherwise the disjunction is classified as inclusive. The tree also finds conjunctive interpretations of disjunction more likely in declarative sentences than interrogatives.

Figure 21 shows the average F1 score of the conjunctive interpretations (AND) for the baseline and the cue-based models. Since the vast majority of the conjunctive interpretations are predicted by the presence of the word *and*, the baseline and cue-based models show similar performances. Setting aside conjunction examples, Figure 22 shows the average F1 score of the AND interpretation of disjunction only. Here we see that the cue-based model performs better than the default model in guessing conjunctive interpretations of disjunction. The informal analysis of the trees suggest that the model does this by using the "speech act" cue. Figure 23 shows the average F1-score of the exclusive interpretations (XOR) for the baseline and the cue-based models. The cue-based model does slightly better than the baseline model. As before, the most important improvement comes in identifying inclusive examples. Figure 24 shows the average F1-score of the inclusive interpretations (IOR) for both baseline and cue-based models. The baseline model performs very poorly while the cue-based model is capable of classifying inclusive examples as well.

Finally, welook at decision trees trained on the annotation data to predict all the interpretation classes for disjunction: AND, XOR, IOR, NOR, and NPQ. Figure 25 shows the baseline model that only uses the words *and* and *or* to classify. As expected, *and* receives a conjunctive interpretation (AND) and *or* receives an exclusive interpretation (XOR). Figure 26 shows the best example tree of the cue-based model. The leaves of the tree show that it recognizes exclusive, inclusive, conjunctive, and even negative inclusive (NOR) interpretations of disjunction. How does the tree achieve that? Like the baseline model, the tree first asks about the connective used: *and* vs. *or*. Then like the previous models, it asks about intonation and consistency. If the intonation is rise-fall, or the disjuncts are inconsistent, the interpretation is exclusive. Then it asks whether the sentence is an interrogative or a declarative. If interrogative, it guesses an inclusive interpretation. This basically covers questions with a rising intonation. Then the tree picks declarative examples that have conditional speech act (e.g. “give me the toy or you’re grounded”) and labels them as exclusive. Finally, if negation is present in the sentence, the tree labels the disjunction as NOR.

Figures 27, 28, and 29 show the average F1-scores for the conjunctive (AND), exclusive (XOR), and inclusive (IOR) interpretations as a function of training size. The results are similar to what wereported before with the ternary classification. While the cue-based model generally performs better than the baseline model, it shows substantial improvement in classifying inclusive cases.

Figure 30 shows the average F1-score for the negative inclusive interpretation as a function of training size. Compared to the baseline model, the cue-based model shows a substantially better performance in classifying negative sentences. The success of the model in classifying negative inclusive examples (NOR) suggests that the cue-based model offers a promising approach for capturing the scope relation of operators such as negation and disjunction. Here, the model learns that when negation and disjunction are present, the sentence receives a negative inclusive (NOR) interpretation. In other words, the model has learned the narrow-scope interpretation of negation and disjunction from the input data. In a language where negation and disjunction receive an XOR interpretation (not A or not B), the cue-based model can learn the wide-scope interpretation of disjunction.

Finally, Figure 31 shows the average F1 score for the class NPQ. This interpretation suggested that the first disjunct is false but the second true. It was seen in examples of repair most often and the most likely cue to it was also the communicative function or speech act of repair. The results show that even though there were improvements in the cue-based model, they were not stable as shown by the large confidence intervals. It is possible that with larger training samples, the cue-based model can reliably classify the NPQ interpretations as well.

Discussion

We considered two accounts for the acquisition of function words. The first account was a baseline (context-independent) account that is used in vanilla cross-situational word learning: words are isolated and directly mapped to their most frequent meanings. The second account is what I called the cue-based context-dependent mapping in which words are mapped to meanings conditional on a set of present cues in the context. I argued that the puzzle of learning disjunction arises because in the baseline account, forms are mapped directly to meanings without considering the context of use. Under this account, the input statistics supports an exclusive interpretation for *or*. However, comprehension studies show that children can interpret *or* as inclusive. I showed that the cue-based account resolves this problem by allowing *or* to be mapped to its interpretation according to the set of contextual cues that disambiguate it. The results of computational experiments with decision tree learning on data from child-directed speech suggested that such an approach can successfully learn to classify a disjunction is inclusive or exclusive. More broadly, cue-based context-dependent mapping is useful for the acquisition of ambiguous words and interpretations that are consistent but relatively infrequent in child-directed speech.

Conclusion

The case of disjunction shows that word learning requires to systmatically take different aspects of the linguistic and non-linguistic context into account. The meaning of a word such as *or* cannot be learned independent of its context such as its intonation contour, the meaning of the coordinands it conjoins, or type of speech act it participates in.

References

Appendix

Properties of CHILDES Corpora

In this section, I report some results on the distribution of words and utterances among the speakers in our collection of corpora. The collection contained 14,159,609 words. Table (1) shows the total number of *and*'s, *or*'s, and words in the speech of children, fathers, and mothers. The collection contains 9 times more words for mothers compared to fathers and 2 more words for mothers compared to children. Therefore, the collection is more representative of the mother-child interactions than father-child interactions. Compared to *or*, the word *and* is 10.80 times more likely in the speech of mothers, 9.20 times more likely in the speech of fathers, and 30.30 times more likely in the speech of children. Overall, *and* is 13.35 times more likely than *or* in this collection which is close to the rate reported by Morris (2008) who used a smaller subset of CHILDES. He extracted 5,994 instances of *and* and 465 instances of *or* and found that overall, *and* was 12.89 times more frequent than *or* in parent-child interactions.

Table 1

Number of and's, or's, and the total number of words in the speech of children and their parents in English-North America and English-UK collections after exclusions.

Speaker Role	and	or	total
Father	15,488	1,683	967,075
Mother	153,781	14,288	8,511,478
Target_Child	78,443	2,590	4,681,056

Figure **??** shows the number of words spoken by parents and children at each month of the child's development. The words in the collection are not distributed uniformly and there is a high concentration of data between the ages of 20 and 40 months (around 2 to 3 years of age). There is also a high concentration around 60 months (5 years of age). The speech of fathers shows a relatively low word-count across all ages. Therefore, in our analyses we should be more cautious in drawing conclusions about the speech of fathers generally, and the speech of mothers and children after age 5. The distribution of function words is sensitive to the type of utterance or more broadly the type of speech act produced by speakers. For example, it is not surprising to hear a parent say "go to your room" but a child saying the same to a parent is unexpected. If a function word commonly occurs in such speech acts, it is unlikely to be produced by children, even though they may understand it very well. Therefore, it is important to check the distribution of speech acts in corpora when studying different function words. Since it is hard to classify and quantify speech acts automatically, here I use utterance type as a proxy for speech acts. I investigate the distribution of declaratives, questions, and imperatives in this collection of corpora on parent-child interactions. Figure 33 shows the distribution of different utterance types in the speech of parents and children. Overall, most utterances are either declaratives or questions, and there are more declaratives than questions in this collection. While mothers and fathers show similar proportions of declaratives and questions in their speech, children produce a lower proportion of questions and higher proportion of declaratives than their parents.

Figure 34 shows the developmental trend of declaratives and questions between the ages of one and six. Children start with only producing declaratives and add non-declarative utterances to their repertoire gradually until they get closer to the parents' rate around the age six. They also start with very few questions and increase the number of questions they ask gradually. It is important to note that the rates of declaratives and questions in children's speech do not reach the adult rate. These two figures show that parent-child interactions are asymmetric. Parents ask more questions and children produce more declaratives. This asymmetry also interacts with age: the speech of younger children has a higher proportion of declaratives than older children.

The frequency of function words such as *and* and *or* may be affected by such conversational asymmetries if they are more likely to appear in some utterance types than others. Figure 35 shows the proportion of *and*'s and *or*'s that appear in different utterance types in parents' and children's speech. In parents' speech, *and* appears more often in declaratives (around 60% in declaratives and 20% in questions). On the other hand, *or* appears more often in questions than declaratives, although this difference is small in mothers. In children's speech, both *and* and *or* appear most often in declaratives. However, children have a higher proportion of *or* in questions than *and* in questions.

The differences in the distribution of utterance types can affect our interpretation of the corpus data on function words such as *and* and *or* in three ways. First, since the collection contains more declaratives than questions, it may reflect the frequency and diversity of function words like *and* that appear in declaratives better. Second, since children produce more declaratives and fewer questions than parents, we may underestimate children's knowledge of function words like *or* that are frequent in questions. Third, given that the percentage of questions in the speech of children increases as they get older, function words like *or* that are more likely to appear in questions may appear infrequent in the early stages and more frequent in the later stages of children's development. In other words, function words like *or* that are common in questions may show a seeming delay in production which is possibly due to the development of questions in children's speech. Therefore, in studying children's productions of function words, it is important to look at their relative frequencies in different utterance types as well as the overall trends. This is the approach I pursue in the next section.

Table 2

Information on the participants in the Providence Corpus. Ethan was diagnosed with Asperger's syndrome and therefore was excluded from this study.

Name	Age Range	Sessions
Alex	1;04.28-3;05.16	51
Ethan	0;11.04-2;11.01	50
Lily	1;01.02-4;00.02	80
Naima	0;11.27-3;10.10	88
Violet	1;02.00-3;11.24	51
William	1;04.12-3;04.18	44

Annotation Categories

Table 3

Annotation classes for connective interpretation

Class	Meaning	Examples
AND	Both propositions are true	*"I'm just gonna empty this and then I'll be out of the kitchen." – "I'll mix them together or I could mix it with carrot, too."*
IOR	One or both propositions are true	*"You should use a spoon or a fork." – "Ask a grownup for some juice or water or soy milk."*
XOR	Only one proposition is true	*"Is that a hyena? or a leopard?" – "We're gonna do things one way or the other."*

Class	Meaning	Examples
NOR	Neither proposition is true	*“I wouldn’t say boo to one goose or three.” – “She found she lacked talent for hiding in trees, for chirping like crickets, or humming like bees.”*
IFF	Either both propositions are true or both are false	*“Put them [crayons] up here and you can get down. – Come over here and I’ll show you.”*
NAB	The first proposition is false, the second is true.	*“There’s an Oatio here, or actually, there’s a wheat here.”*

Table 4

Definitions of the intonation types and their examples.

Intonation	Definitions	Examples
Flat	Intonation does not show any substantial rise at the end of the sentence.	*“I don’t hear any meows or bow-wow-wows.”*
Rise	There is a substantial intonation rise on each disjunct or generally on both.	*“Do you want some seaweed? or some wheat germ?”*
Rise-Fall	There is a substantial rise on the non-final disjunct(s), and a fall on the final disjunct.	*“Is that big Q or little q?” – “(are) You patting them, petting them, or slapping them?”*

Table 5

Definitions of the utterance types and their examples.

Utterance Types	Definitions	Examples
Declarative	A statement with a subject-verb-object word order and a flat intonation.	*"It looks a little bit like a drum stick or a mallet."*
Interrogative	A question with either subject-auxiliary inversion or a rising terminal intonation.	*"Is that a dog or a cat?"*
Imperative	A directive with an uninflected verb and no subject	*"Have a little more French toast or have some of your juice."*

Table 6

Definitions of the syntactic levels and their examples.

Syntactic Level	Definitions	Examples
Clausal	The coordinands are sentences, clauses, verb phrases, or verbs.	*"Does he lose his tail sometimes and Pooh helps him and puts it back on?"*
Sub-clausal	The coordinands are nouns, adjectives, noun phrases, determiner phrases, or prepositional phrases.	*"Hollies can be bushes or trees."*

Table 7

Definitions of consistency types and their examples.

Consistency	Definitions	Examples
Consistent	The coordinands can be true at the same time.	*"We could spell some things with a pen or draw some pictures."*
Inconsistent	The coordinands cannot be true at the same time.	*"Do you want to stay or go?"*

Table 8

Definitions of the communicative functions and their examples.

Function	Definitions	Examples
Descriptions	Describing what the world is like or asking about it. The primary goal is to inform the addressee about how things are.	"*It's not in the ditch or the drain pipe.*"
Identifications	Identifying the category membership or an attribute of an object. Speaker has uncertainty. A subtype of "Description".	"*Is that a ball or a balloon honey?*"
Definitions and Examples	Providing labels for a category or examples for it. Speaker is certain. Subtype of Description.	*"This is a cup or a mug." – "berries like blueberry or raspberry"*
Preferences	Asking what the addressee wants or would like or stating what the speaker wants or would like	*"Do you wanna play pizza or read the book?"*

Function	Definitions	Examples
Options	Either asking or listing what one can or is allowed to do. Giving permission, asking for permission, or describing the possibilities. Often the modal "can" is either present or can be inserted.	*"You could have wheat or rice."*
Directives	Directing the addressee to act or not act in a particular way. Common patterns include "let's do ...", "Why don't you do ...", or prohibitions such as "Don't ...". The difference with "options" is that the speaker expects the directive to be carried out by the addressee. There is no such expectation for "options".	*"let's go back and play with your ball or we'll read your book."*
Clarifications	Something is said or done as a communicative act but the speaker has uncertainty with respect to the form or the content.	*"You mean boba or bubble?"*
Repairs	Speaker correcting herself on something she said (self repair) or correcting the addressee (other repair). The second disjunct is what holds and is intended by the speaker. The speaker does not have uncertainty with respect to what actually holds.	*"There's an Oatio here, or actually, there's a wheat here."*

Function	Definitions	Examples
Conditionals	Explaining in the second coordinand, what would follow if the first coordinand is (or is not) followed. Subtype of Directive.	*"Put that out of your mouth, or I'm gonna put it away."* – *"Come over here and I'll show you."*
Unconditionals	Denying the dependence of something on a set of conditions. Typical format: "Whether X or Y, Z". Subtype of Descriptions.	*"Ready or not, here I come!"* (playing hide and seek)

Table 9

Definitions of answer types and their examples.

Type	Definitions	Examples
No Answer	The child provides no answer to the question.	Mother: *"Would you like to eat some applesauce or some carrots?"* Child: *"Guess what Max!"*
YN	The child responds with *yes* or *no*.	Father: *"Can I finish eating one or two more bites of my cereal?"* Child: *"No."*
AB	The child responds with one of the disjuncts (alternatives).	Mother: *"Is she a baby elephant or is she a toddler elephant?"* Child: *"It's a baby. She has a tail."*

Inter-annotator agreement

Figure 37 shows the percentage agreement and the kappa values for each annotation category over the 8 iterations.

Agreement in the following three categories showed substantial improvement after better and more precise definitions and annotation criteria were developed: connective interpretation, intonation, and communicative function. First, connective interpretation showed major improvements after annotators developed more precise criteria for selecting the propositions under discussion and separately wrote down the two propositions connected by the connective word. For example, if the original utterance was “do you want milk or juice?”, the annotators wrote “you want milk, you want juice” as the two propositions under discussion. This exercise clarified the exact propositions under discussion and sharpened annotator intuitions with respect to the connective interpretation that is communicated by the utterance. Second, annotators improved agreement on intonation by reconstructing an utterance’s intonation for all three intonation categories. For example, the annotator would examine the same sentence “do you want coffee or tea?” with a rise-fall, a rise, and a flat intonation. Then the annotator would listen to the actual utterance and see which one most resembled the actual utterance. This method helped annotators judge the intonation of an utterance more accurately. Finally, agreement on communicative functions improved as the definitions were made more precise. For example, the definition of “directives” in Table 8 explicitly mentions the difference between “directives” and “options”. Clarifying the definitions of communicative functions helped improve annotator agreement.

Inter-annotator reliability for conjunction was calculated in the same way. Two different annotators coded 300 utterances of *and*. Inter-annotator reliability was calculated over 10 iterations of 30 examples. Figure 38 shows the percentage agreement between the annotators as well as the kappa values for each iteration. Despite high percentage agreement between annotators, the kappa values did not pass the set threshold of 0.7 in three consecutive iterations. This paradoxical result is mainly due to a property of kappa. An imbalance in the prevalence of annotation categories can drastically lower its value. When one category is extremely common with high agreement while other categories are rare, kappa will be low (Cicchetti & Feinstein, 1990; Feinstein & Cicchetti, 1990). In almost all annotated categories for conjunction, there was one class that was extremely prevalent. In such cases, it is more informative to look at the class specific agreement for the prevalent category than the overall agreement measured by Kappa (Cicchetti & Feinstein, 1990; Feinstein & Cicchetti, 1990).

Table 10 lists the dominant classes as well as their prevalence, the values of class specific agreement index, and category agreement index (Kappa). Class specific agreement index is defined as $2n_{ii}/n_{i.} + n_{.i}$, where i represents the class's row/column number in the category's confusion matrix, n the number of annotations in a cell, and the dot ranges over all the row/column numbers (Fleiss, Levin, & Paik, 2013, p. 600; Ubersax, 2009). The class specific agreement indices are high for all the most prevalent classes showing that the annotators had very high agreement on these class, even though the general agreement index (Kappa) was often low. The most extreme case is the category "consistency" where almost all instances were annotated as "consistent" with perfect class specific agreement but low overall Kappa. In the case of utterance type and syntactic level where the distribution of instances across classes was more even, the general index of agreement Kappa is also high. In general, examples of conjunction showed little variability across annotation categories and mostly fell into one class within each category. Annotators had high agreement for these dominant classes.

Braine, M. D., & Rumain, B. (1981). Development of comprehension of "or": Evidence for a sequence of competencies. *Journal of Experimental Child Psychology*, *31*(1), 46–70.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Table 10

Most prevalent annotation class in each annotation category with the values of class agreement indeces and category agreement indeces (Kappa).

Annotation Category	Class	Prevalence	Class Agreement Index	Kappa
intonation	flat	0.86	0.89	0.24
interpretation	AND	0.96	0.98	0.39
answer	NA	0.84	0.94	0.67
utterance_type	declarative	0.76	0.94	0.70
communicative_function	description	0.77	0.90	0.59
syntactic_level	clausal	0.67	0.91	0.70
consistency	consistent	0.99	1.00	0.50

Breiman, L. (2017). *Classification and regression trees*. London: Routledge.

Brown, R. (1957). Linguistic determinism and the part of speech. *The Journal of Abnormal and Social Psychology*, *55*(1), 1.

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In *Proceedings of the 25th Boston University conference on language development* (pp. 157–168). Somerville, MA: Cascadilla Press.

Chierchia, G., Guasti, M. T., Gualmini, A., Meroni, L., Crain, S., & Foppolo, F. (2004). Semantic and pragmatic competence in children's and adults' comprehension of or. In I. Noveck & D. Sperber (Eds.), *Experimental pragmatics* (pp. 283–300). Basingstoke: Palgrave Macmillan.

Cicchetti, D. V., & Feinstein, A. R. (1990). High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology*, *43*(6), 551–558.

Crain, S. (2012). *The emergence of meaning.* Cambridge: Cambridge University Press.

Crain, S., Gualmini, A., & Meroni, L. (2000). The acquisition of logical words. *LOGOS and Language*, *1*, 49–59.

Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech*, *49*(2), 137–173.

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology*, *43*(6), 543–549.

Fisher, C., Gertner, Y., Scott, R. M., & Yuan, S. (2010). Syntactic bootstrapping. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(2), 143–149.

Fleiss, J. L., Levin, B., & Paik, M. C. (2013). *Statistical methods for rates and proportions.* New York: John Wiley & Sons.

Geurts, B. (2006). Exclusive disjunction without implicatures. *Ms., University of Nijmegen.*

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(3), 515–531.

Grice, H. P. (1989). *Studies in the way of words.* Cambridge, MA: Harvard University Press.

Gualmini, A., & Crain, S. (2002). Why no child or adult must learn de Morgan's laws. In *Proceedings of the Boston University conference on language development.* Somerville, MA: Cascadilla Press.

Gualmini, A., Crain, S., & Meroni, L. (2000). Acqisition of disjunction in conditional sentences. In *Proceedings of the boston university conference on language development.*

Haspelmath, M. (2007). Coordination. In T. Shopen (Ed.), *Language typology and linguistic description,* Cambridge: Cambridge University Press.

Ho, T. K. (1995). Random decision forests. In *Proceedings of the third international conference on document analysis and recognition* (Vol. 1, pp. 278–282). Washington, DC, USA: IEEE Computer Society.

Johansson, B. S., & Sjolin, B. (1975). Preschool children's understanding of the coordinators "and" and "or". *Journal of Experimental Child Psychology*, *19*(2), 233–240.

Kamp, H. (1973). Free choice permission. In *Proceedings of the Aristotelian society* (Vol. 74, pp. 57–74).

Levy, E., & Nelson, K. (1994). Words in discourse: A dialectical approach to the acquisition of meaning and use. *Journal of Child Language*, *21*(02), 367–389.

MacWhinney, B. (2000). *The CHILDES project: The database* (Vol. 2). Mahwah, NJ: Erlbaum.

Morris, B. J. (2008). Logically speaking: Evidence for item-based acquisition of the connectives "and" and "or". *Journal of Cognition and Development*, *9*(1), 67–88.

Neisser, U., & Weene, P. (1962). Hierarchies in concept attainment. *Journal of Experimental Psychology*, *64*(6), 640.

Notley, A., Thornton, R., & Crain, S. (2012a). English-speaking children's interpretation of disjunction in the scope of "not every". *Biolinguistics*, *6*(1), 32–69.

Notley, A., Zhou, P., Jensen, B., & Crain, S. (2012b). Children's interpretation of disjunction in the scope of "before": A comparison of English and Mandarin. *Journal of Child Language*, *39*(03), 482–522.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … others. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.

Pruitt, K., & Roelofsen, F. (2013). The interpretation of prosody in disjunctive questions. *Linguistic Inquiry*, *44*(4), 632–650.

Sanchez, A., Meylan, S., Braginsky, M., MacDonald, K., Yurovsky, D., & Frank, M. C. (2018). Childes-db: A flexible and reproducible interface to the child language data exchange system. PsyArXiv. Retrieved from psyarxiv.com/93mwx

Sison, C. P., & Glaz, J. (1995). Simultaneous confidence intervals and sample size determination for multinomial proportions. *Journal of the American Statistical Association*, *90*(429), 366–369.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Harvard University Press.

Ubersax, J. (2009). Retrieved from http://www.john-uebersax.com/stat/raw.htm

Von Wright, G. H. (1968). An essay in deontic logic and the general theory of action.

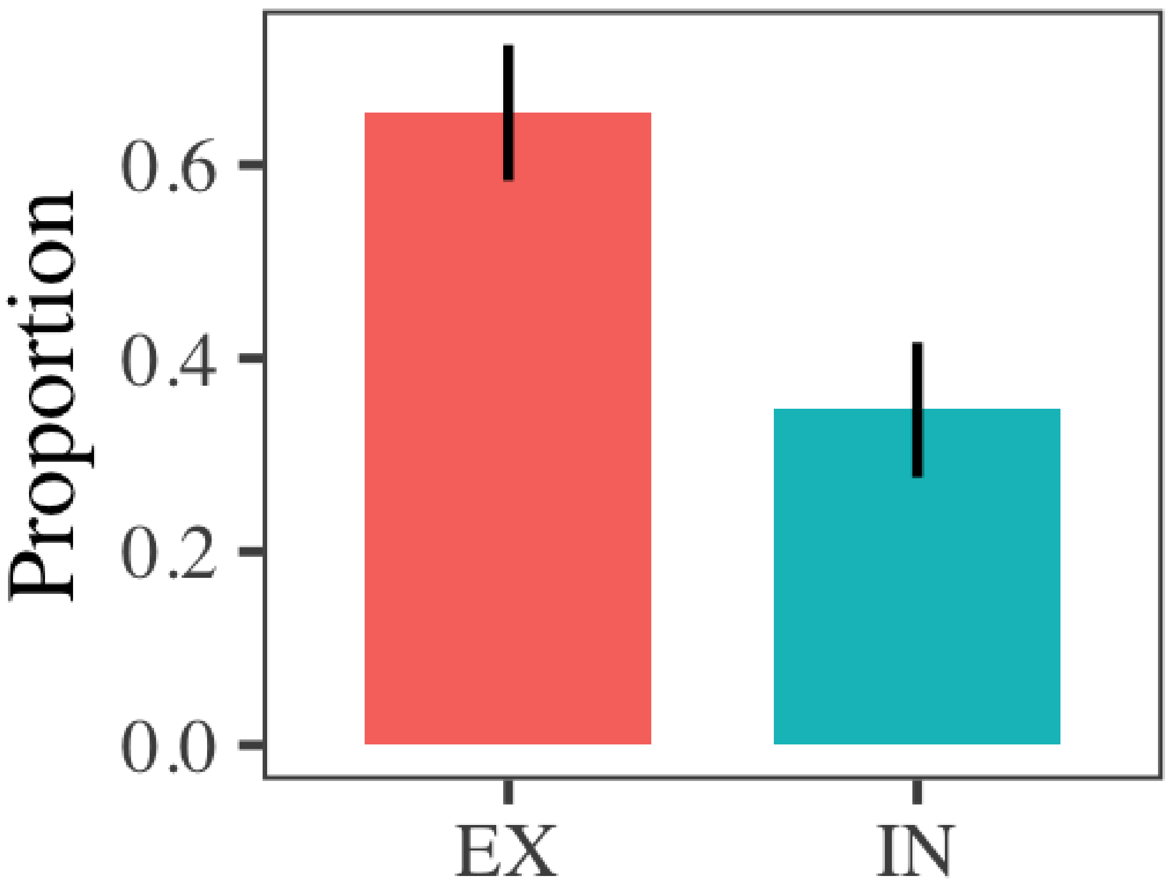


Figure 13. Proportion of exclusive and inclusive interpretations of disjunction in child-directed speech. Error bars represent bootstrapped 95% confidence intervals.

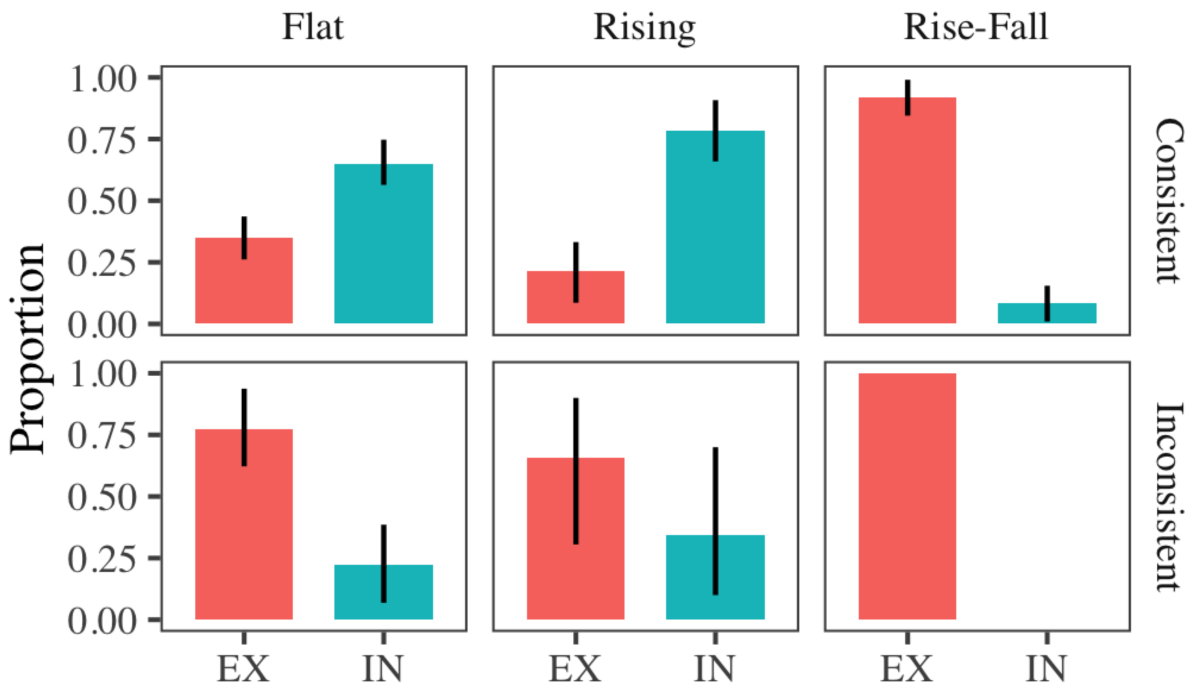


Figure 14. Exclusive and inclusive interpretations broken down by intonation (flat, rise, rise-fall) and consistency. Error bars represent bootstrapped 95% confidence intervals.

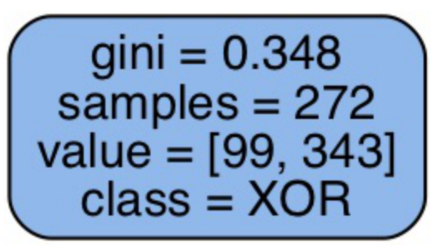


Figure 15. Baseline tree grown with minimum impurity decrease of 0.2. The tree always classifies examples of disjunction as exclusive.

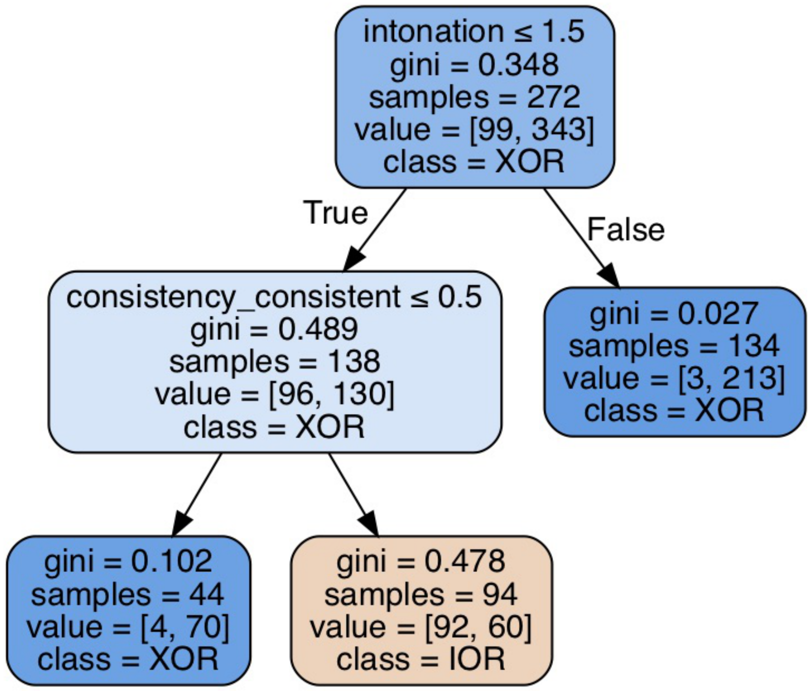


Figure 16. Cue-based tree grown with minimum impurity decrease of 0.01. The tree classifies examples of disjunction with rise-fall intonation as exclusive (intonation > 1.5). If the intonation is not rise-fall but the disjuncts are inconsistent (consistency < 0.5), then the disjunction is still classified as exclusive. However, if neither of these two hold, the disjunction is classified as inclusive.

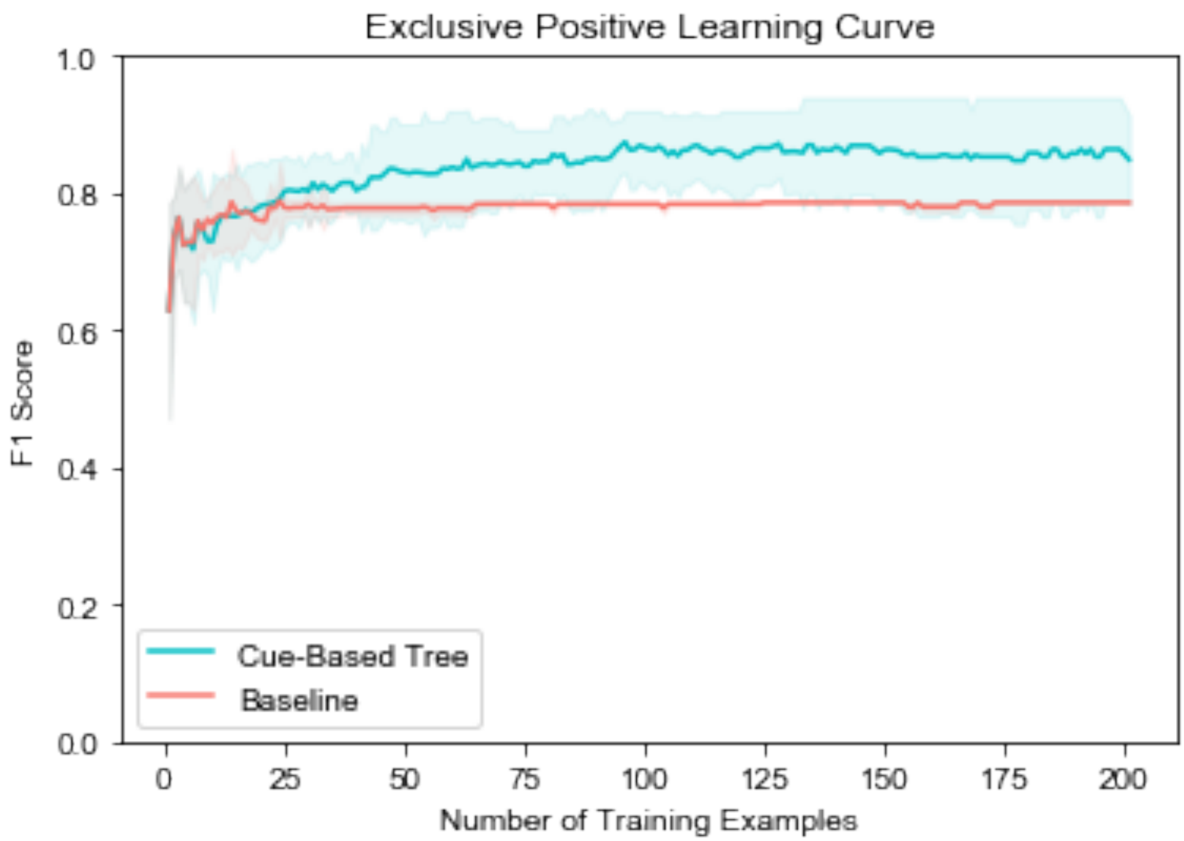


Figure 17. The average F1 score for class XOR (exclusive) as a function of the number of training examples in the baseline and cue-based models. The colored shades show the 95% confidence intervals.

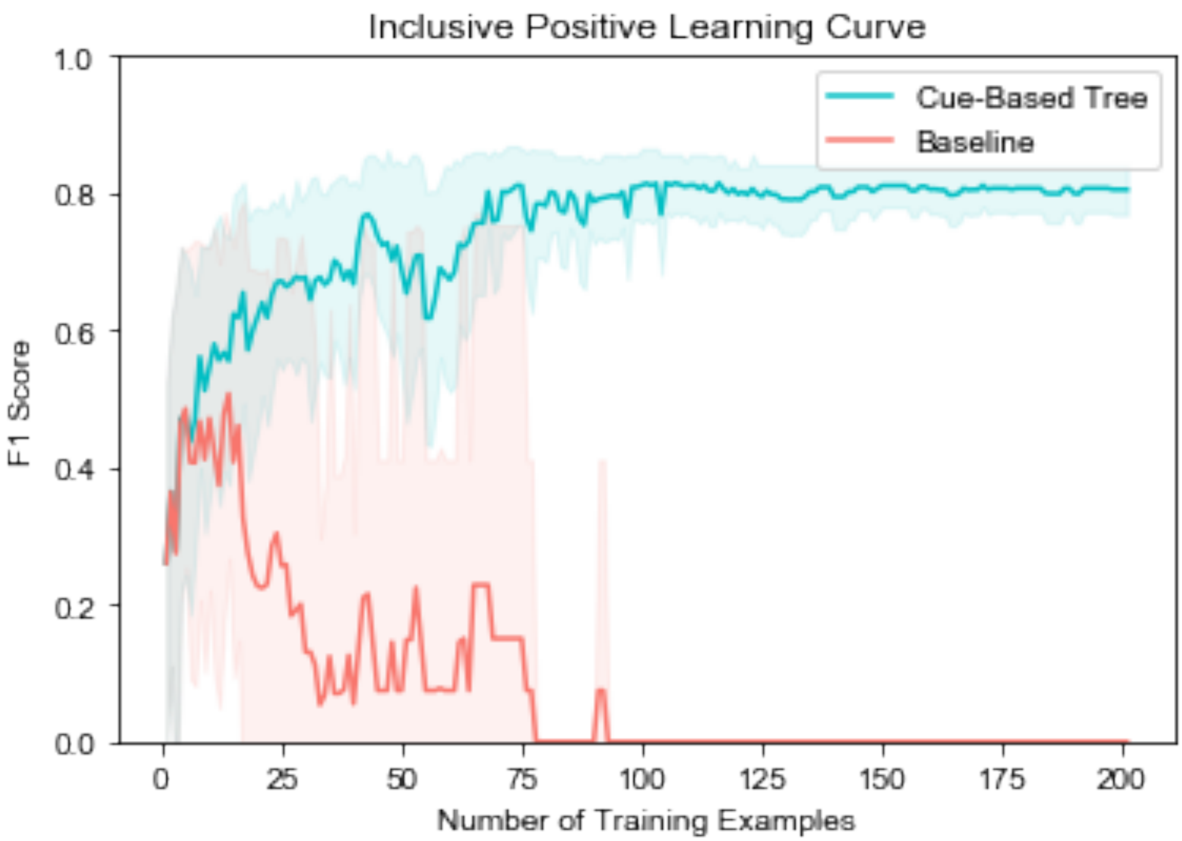


Figure 18. The average F1 score for class IOR (inclusive) as a function of the number of training examples in the baseline and cue-based models. The colored shades show the 95% confidence intervals.

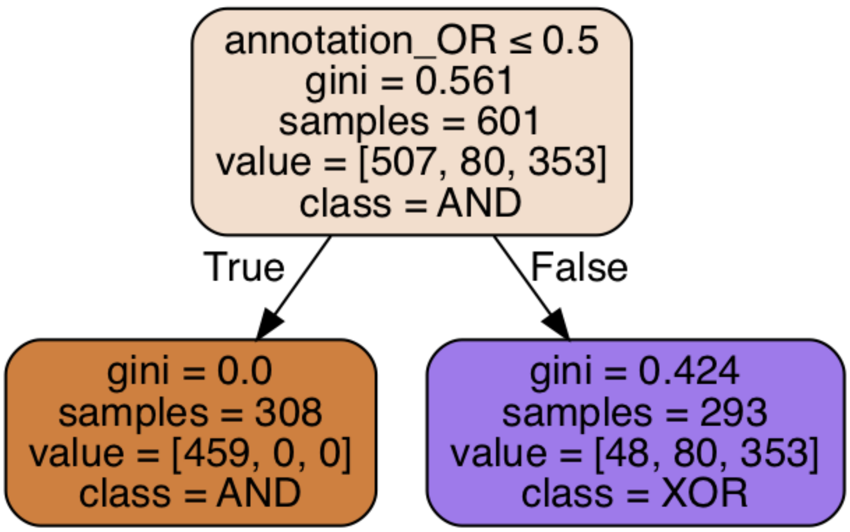


Figure 19. The baseline tree grown on conjunctions and disjunctions with minimum impurity decrease of 0.2. The tree uses the words *and/or* and classifies them as conjunction and exclusive disjunction respectively.

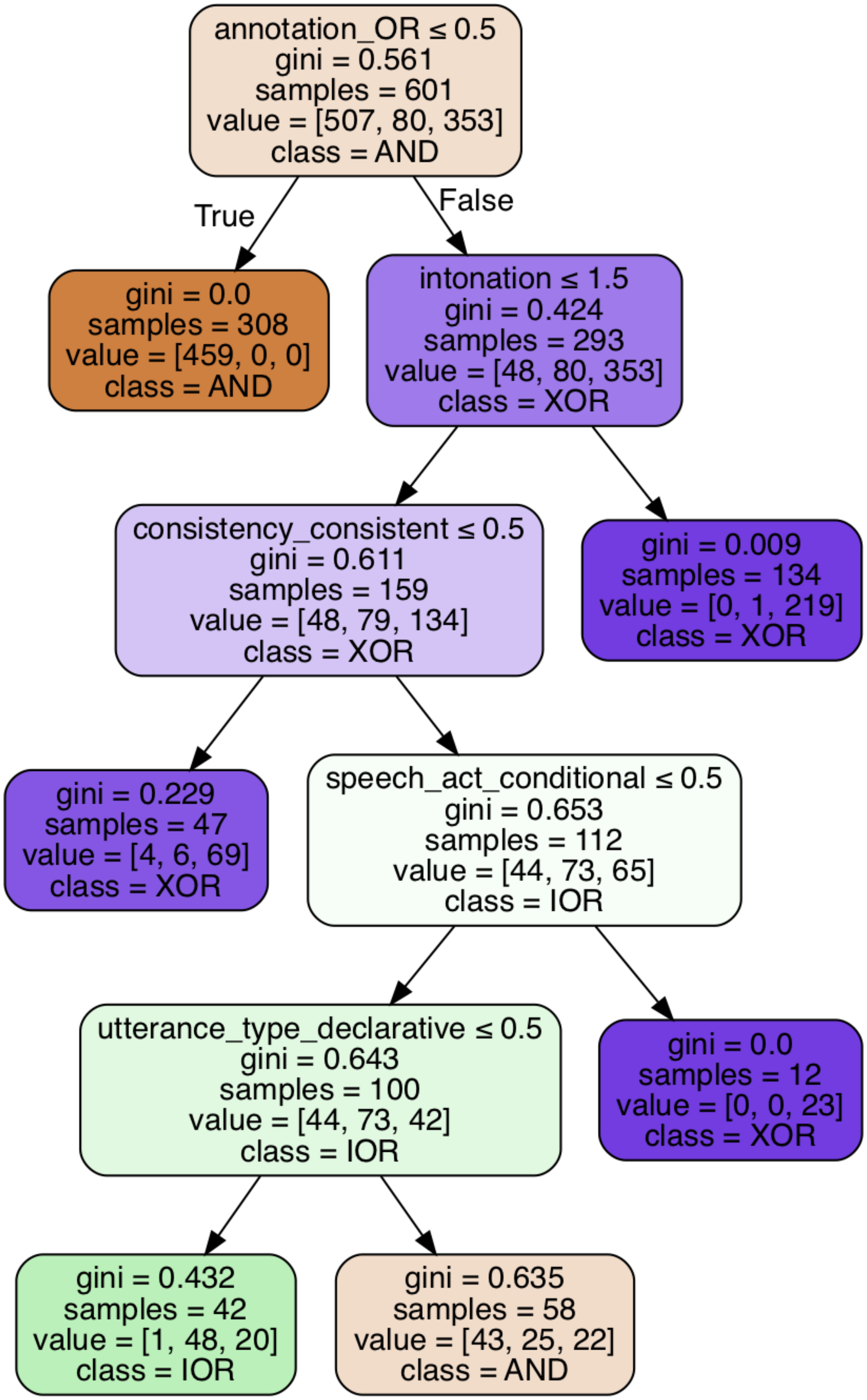


Figure 20. The cue-based tree grown on conjunctions and disjunctions with minimum impurity decrease of 0.01. After using the words *and/or*, the tree uses intonation, consistency,

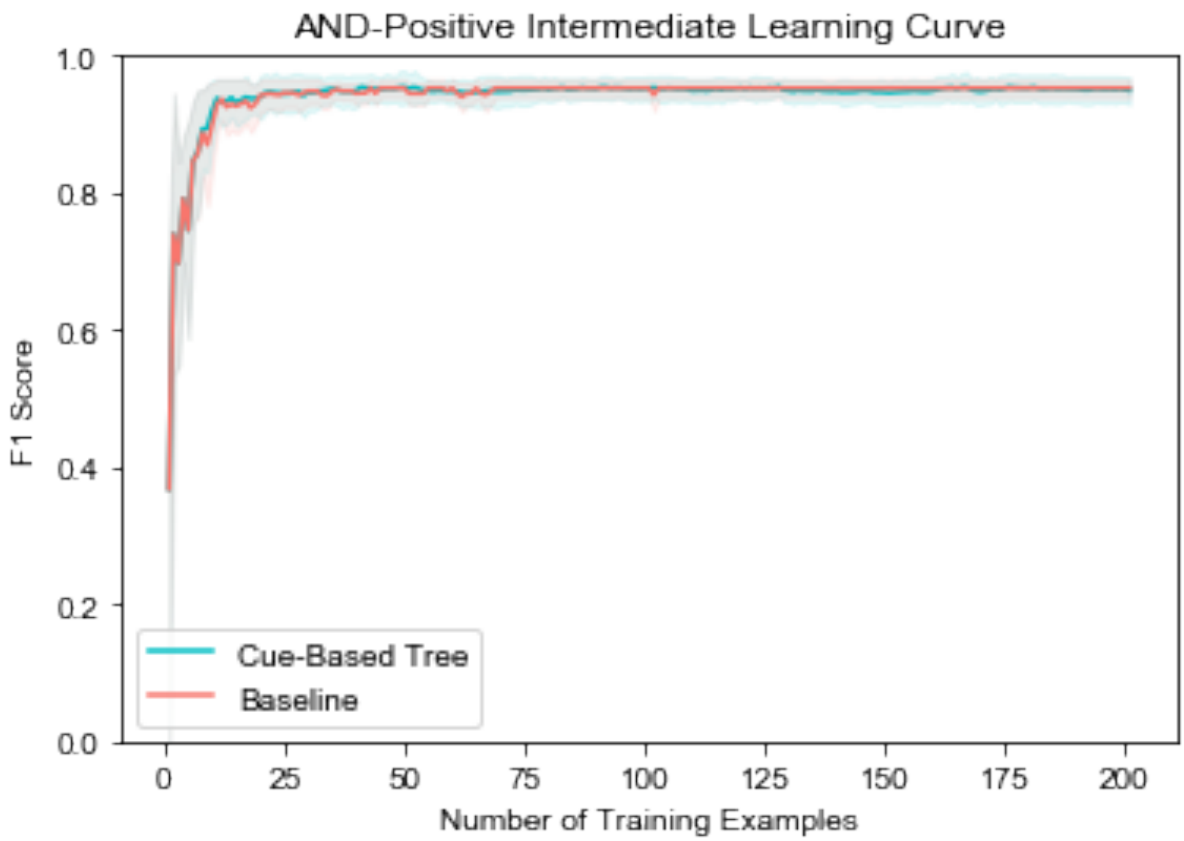


Figure 21. The average F1 score for class AND as a function of the number of training examples in the baseline and cue-based models. The colored shades show the 95% confidence intervals.

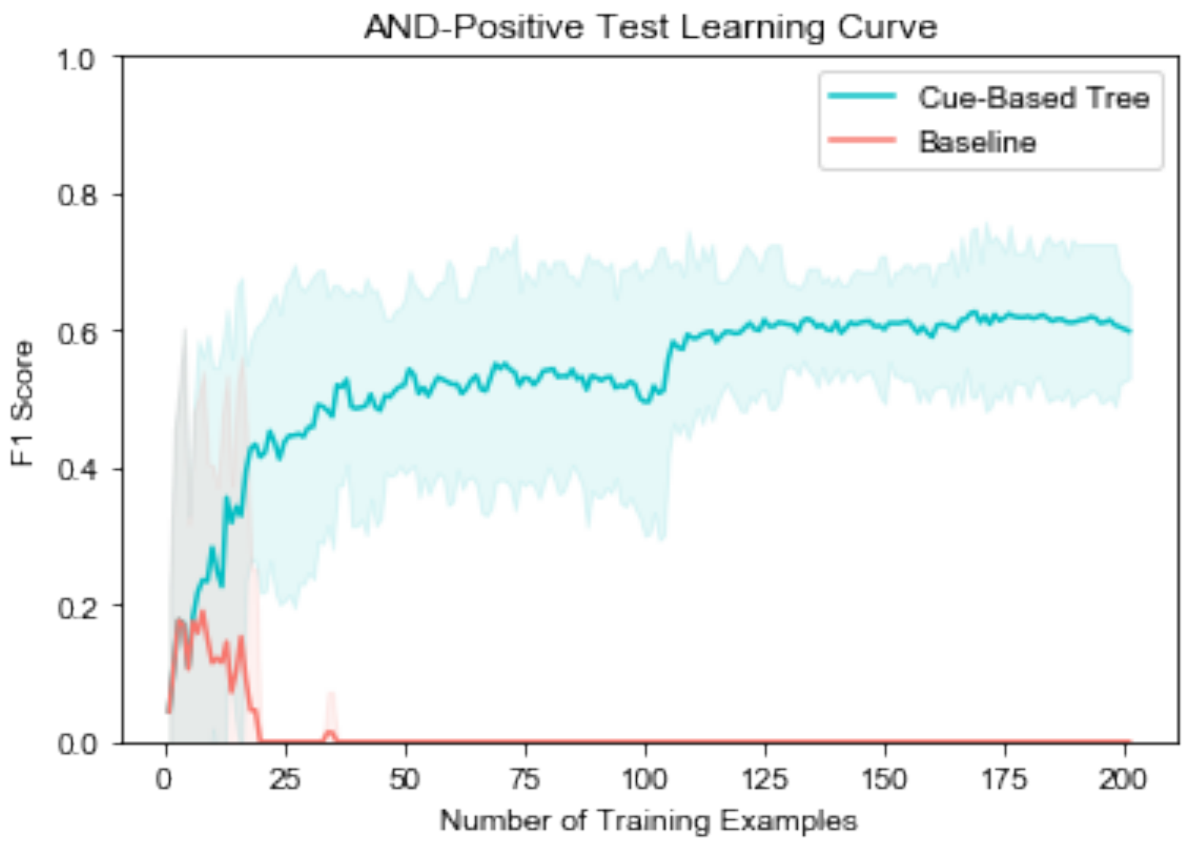


Figure 22. The average F1 score for class AND of disjunction examles as a function of the number of training examples in the baseline and cue-based models. The colored shades show the 95% confidence intervals.

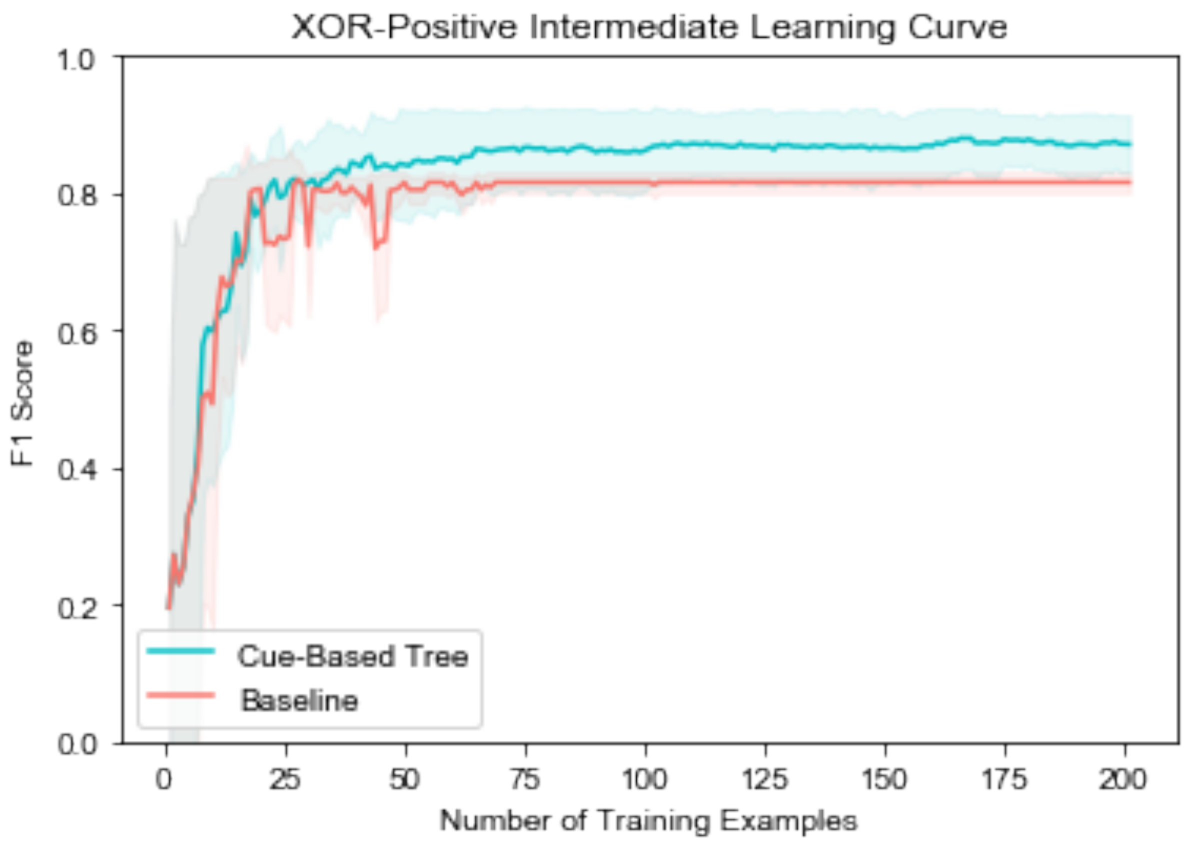


Figure 23. The average F1 score for class XOR as a function of the number of training examples in the baseline and cue-based models. The colored shades show the 95% confidence intervals.

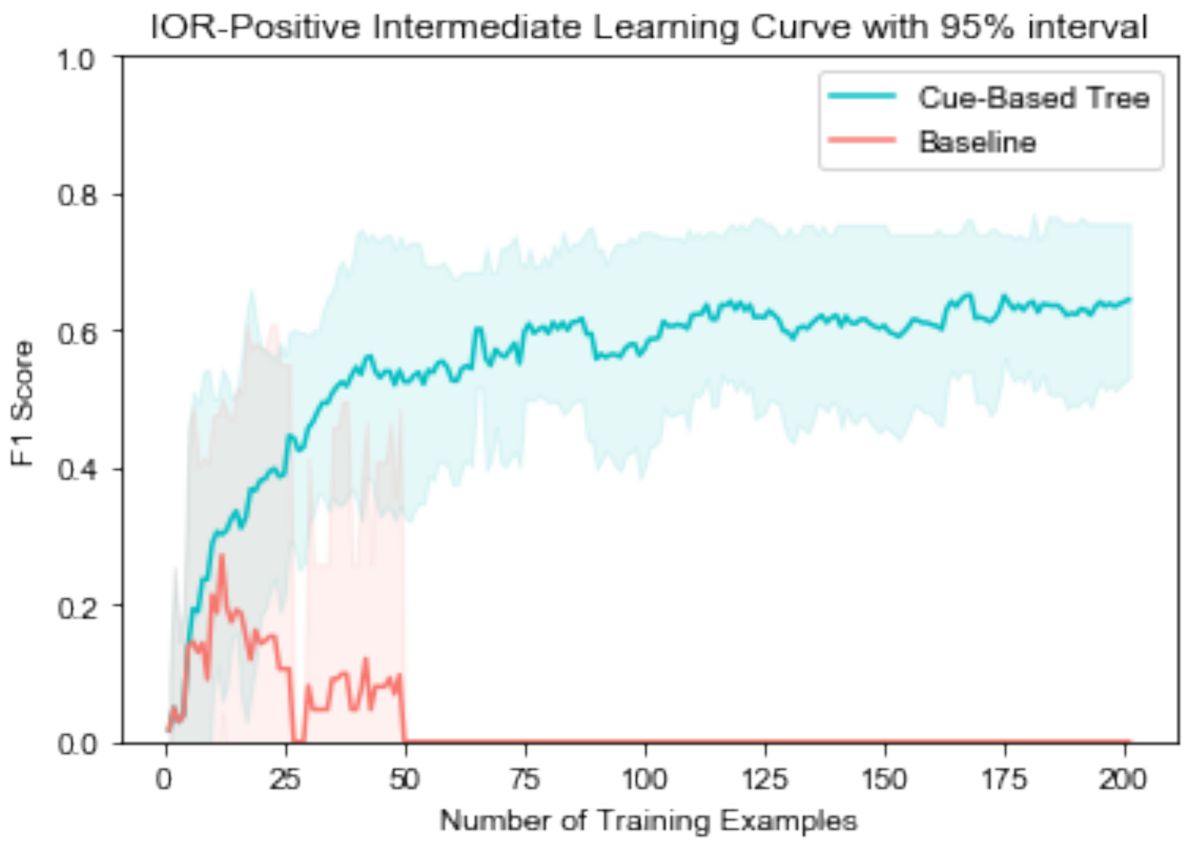


Figure 24. The average F1 score for class IOR as a function of the number of training examples in the baseline and cue-based models. The colored shades show the 95% confidence intervals.

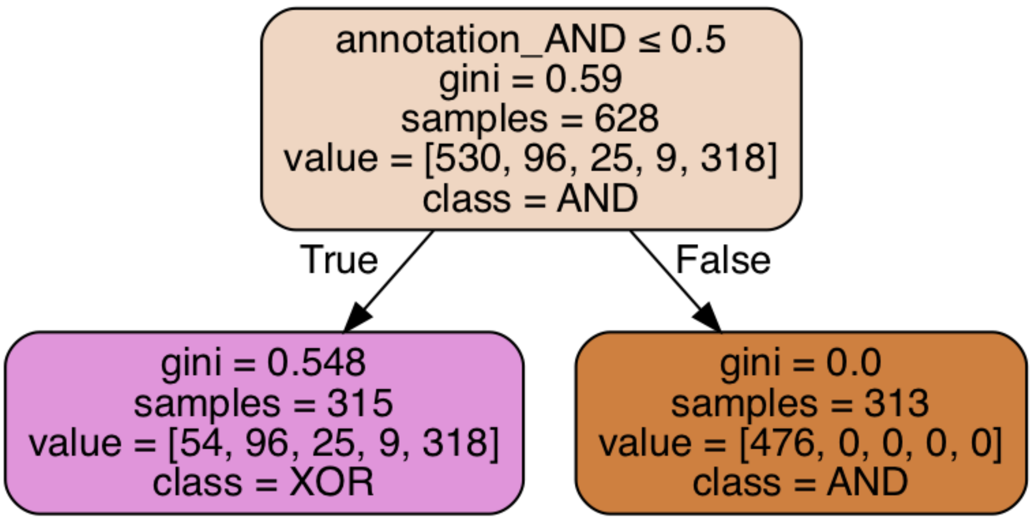


Figure 25. The baseline tree grown on conjunctions and disjunctions with minimum impurity decrease of 0.2. The tree uses the words *and/or* and classifies them as conjunction and exclusive disjunction.

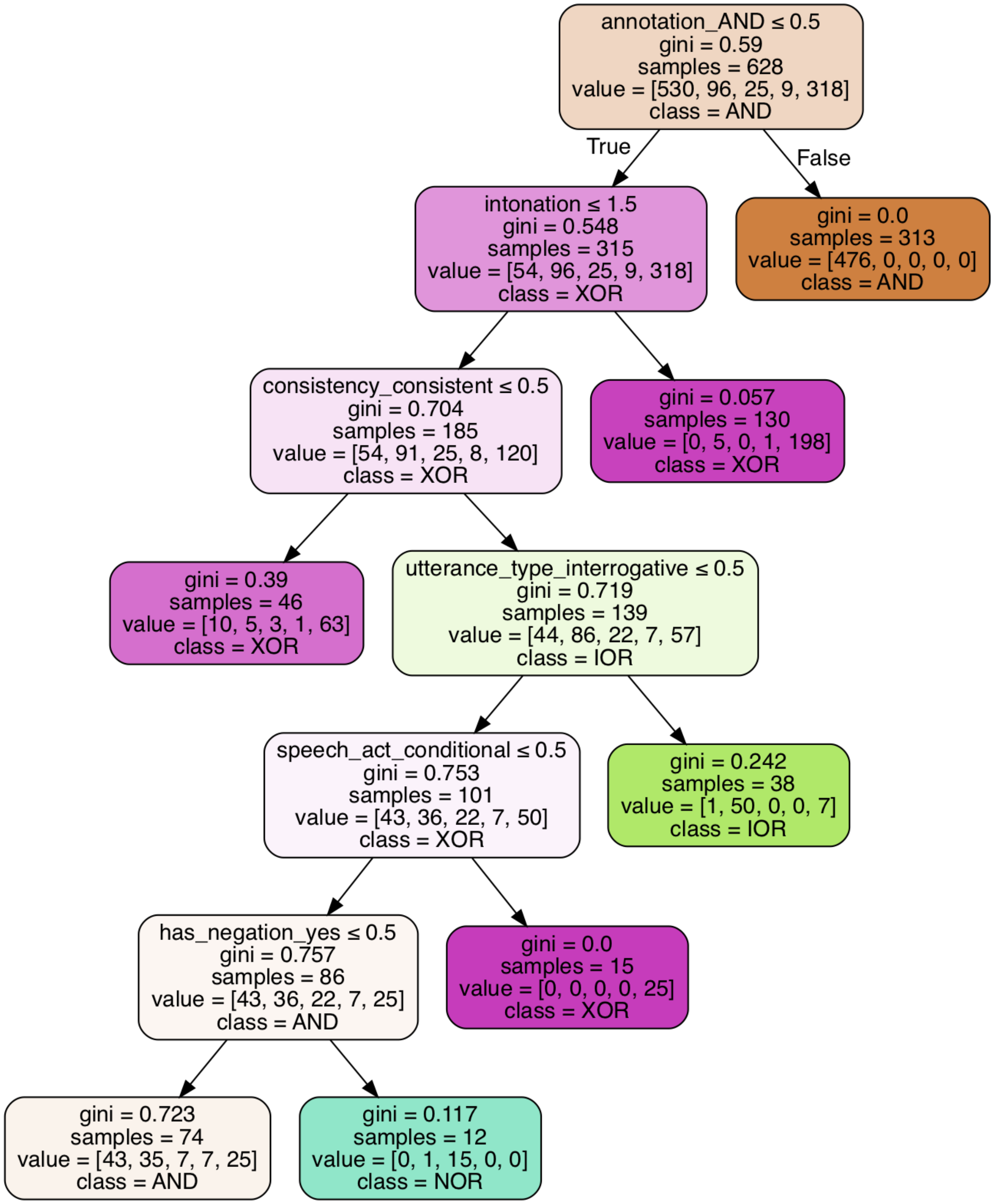


Figure 26. The cue-based tree grown on conjunctions and disjunctions with minimum impurity decrease of 0.01. After using the words *and/or*, the tree uses intonation and consistency to classify a large number of exclusive cases. Then it uses utterance type (interrogative) to label many inclusive cases, as well as the communicative function (conditional) to catch more exclusive examples. Finally, it asks whether the sentence has negation or not. If so, it classifies the negative inlusive examples as NOR.

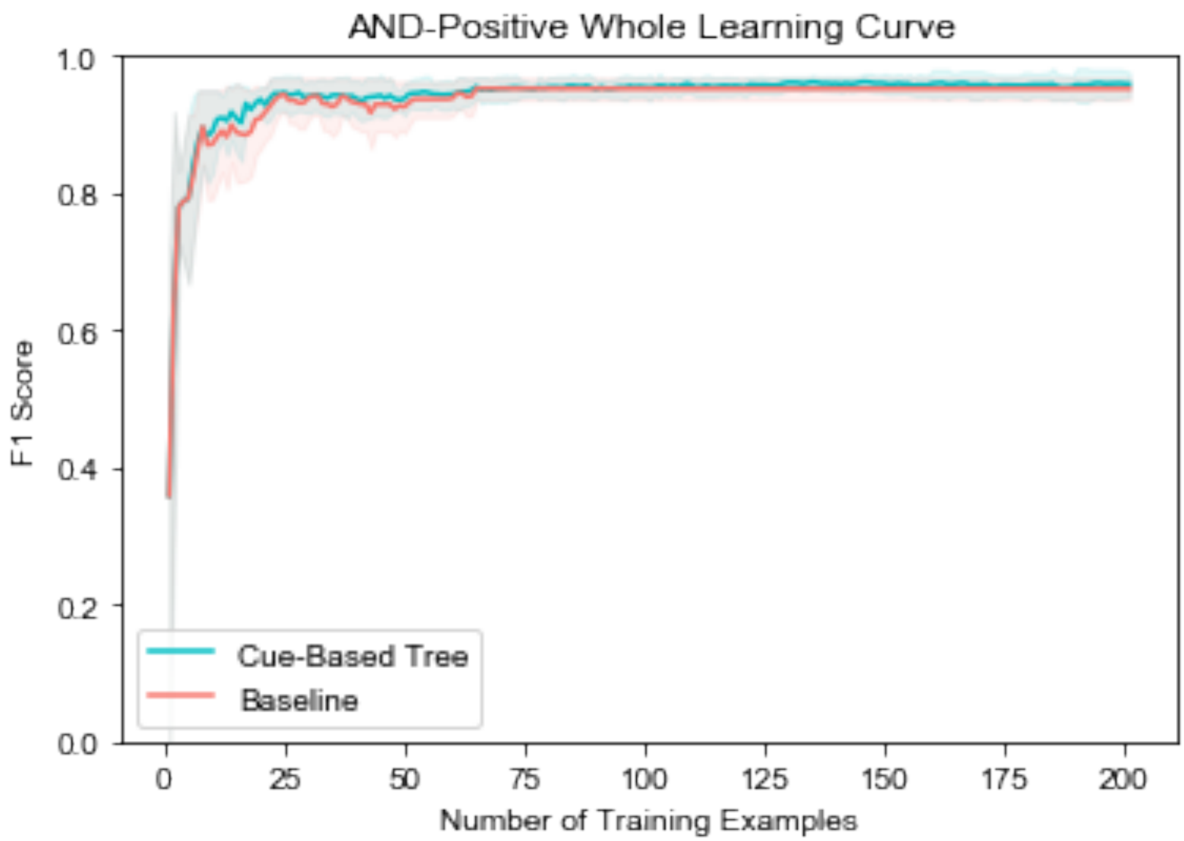


Figure 27. The average F1 score for class AND as a function of the number of training examples in the baseline and cue-based models. The colored shades show the 95% confidence intervals.

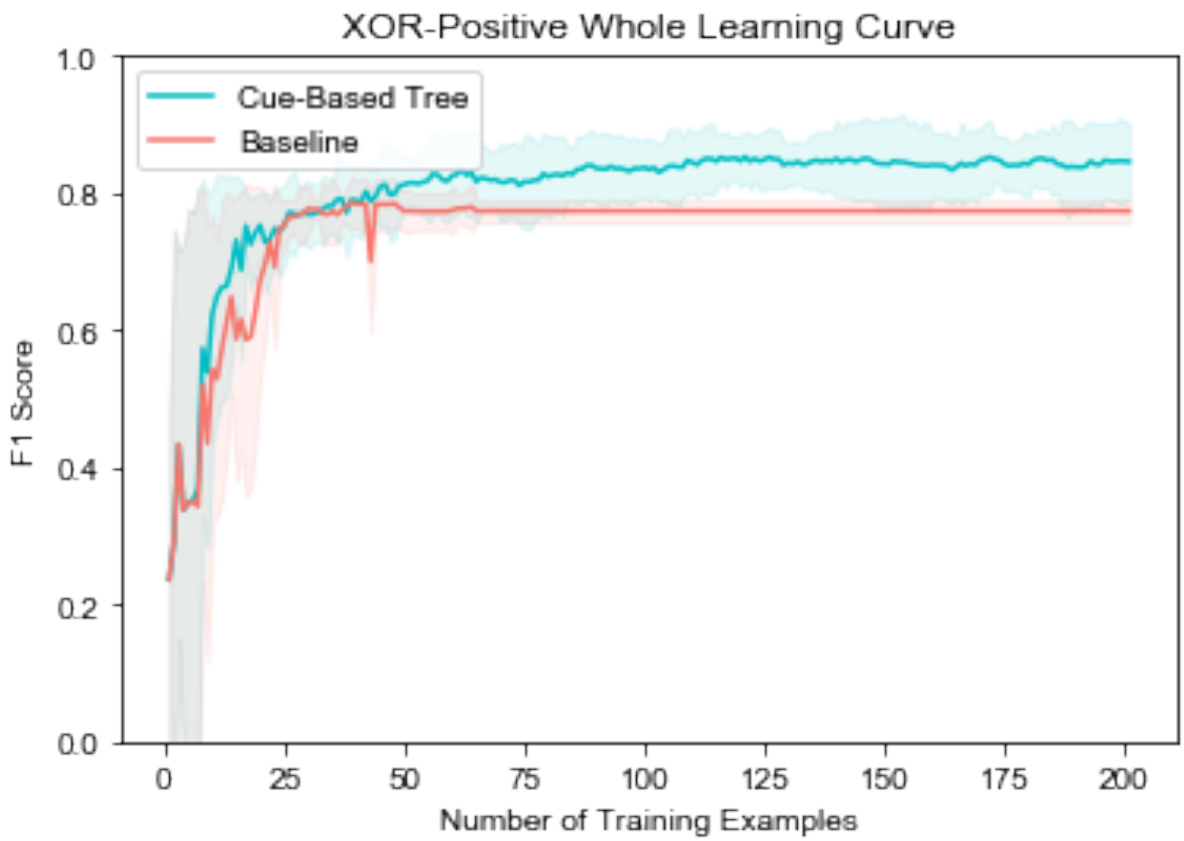


Figure 28. The average F1 score for class XOR as a function of the number of training examples in the baseline and cue-based models. The colored shades show the 95% confidence intervals.

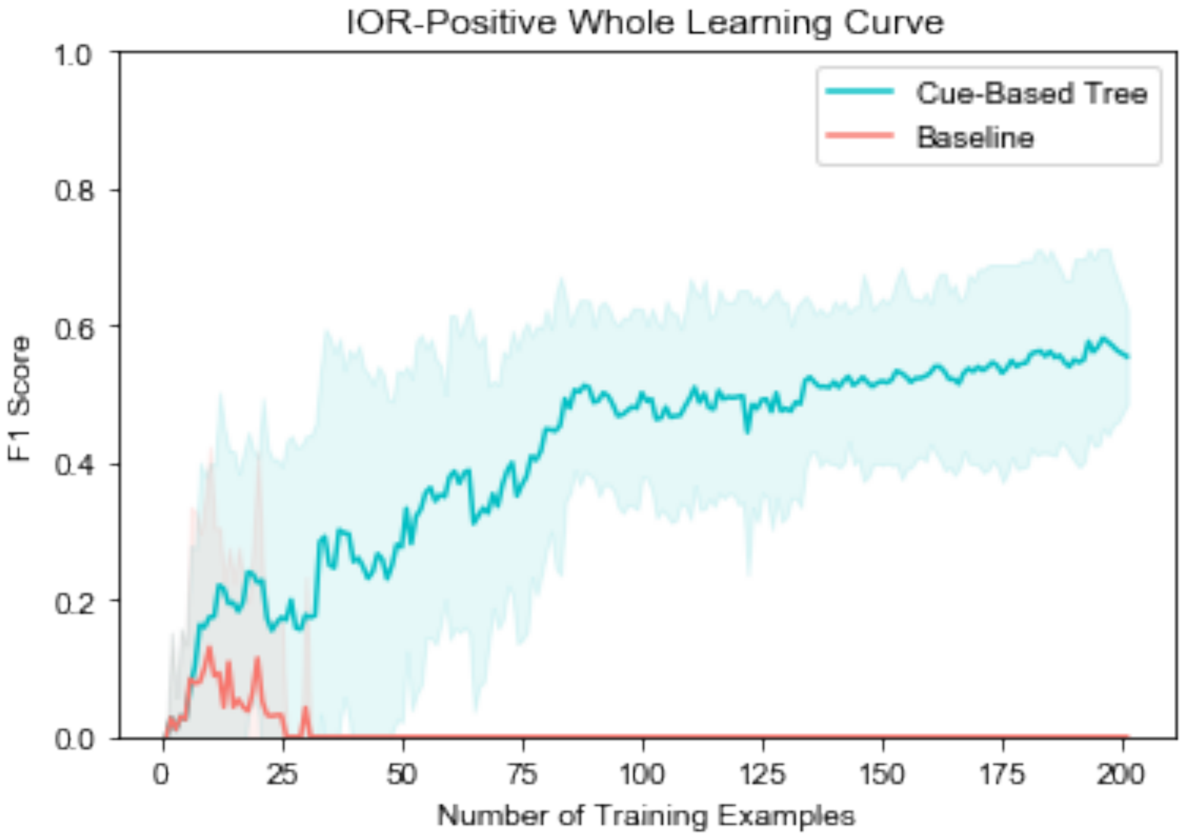


Figure 29. The average F1 score for class IOR as a function of the number of training examples in the baseline and cue-based models. The colored shades show the 95% confidence intervals.

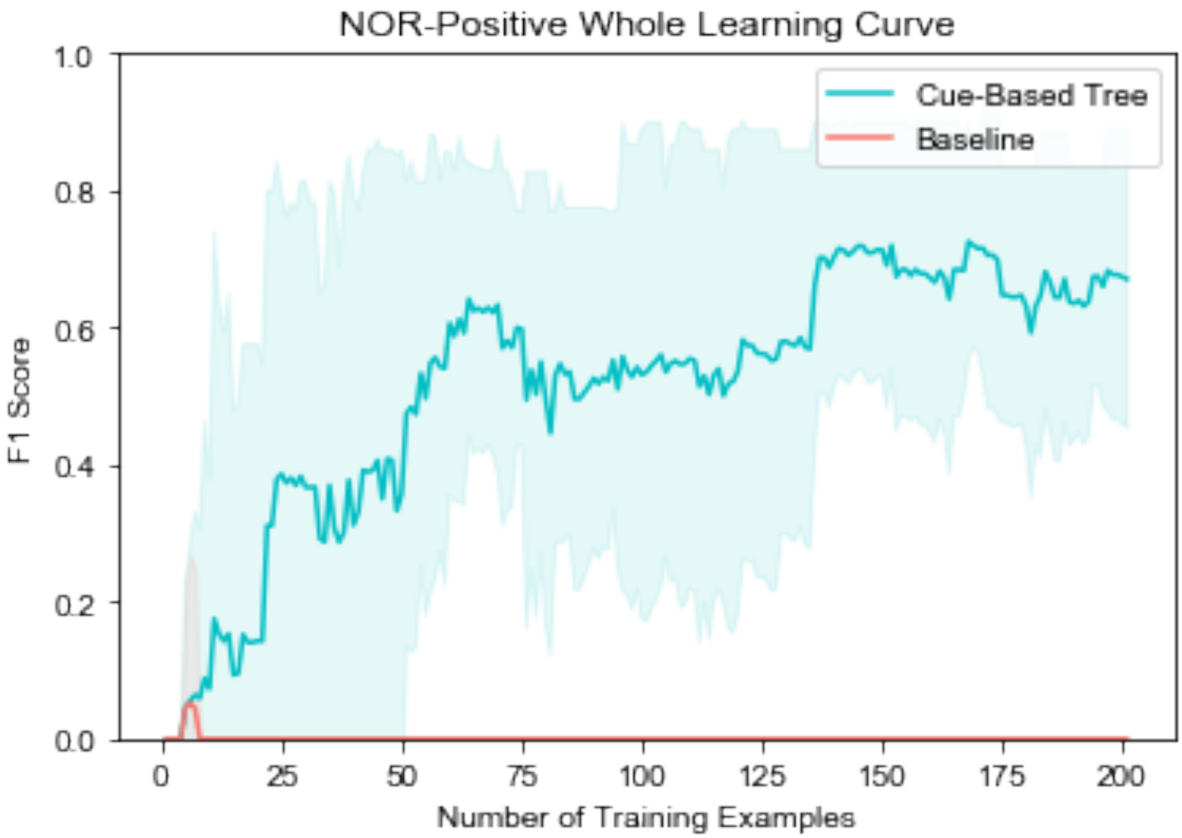


Figure 30. The average F1 score for class NOR as a function of the number of training examples in the baseline and cue-based models. The colored shades show the 95% confidence intervals.

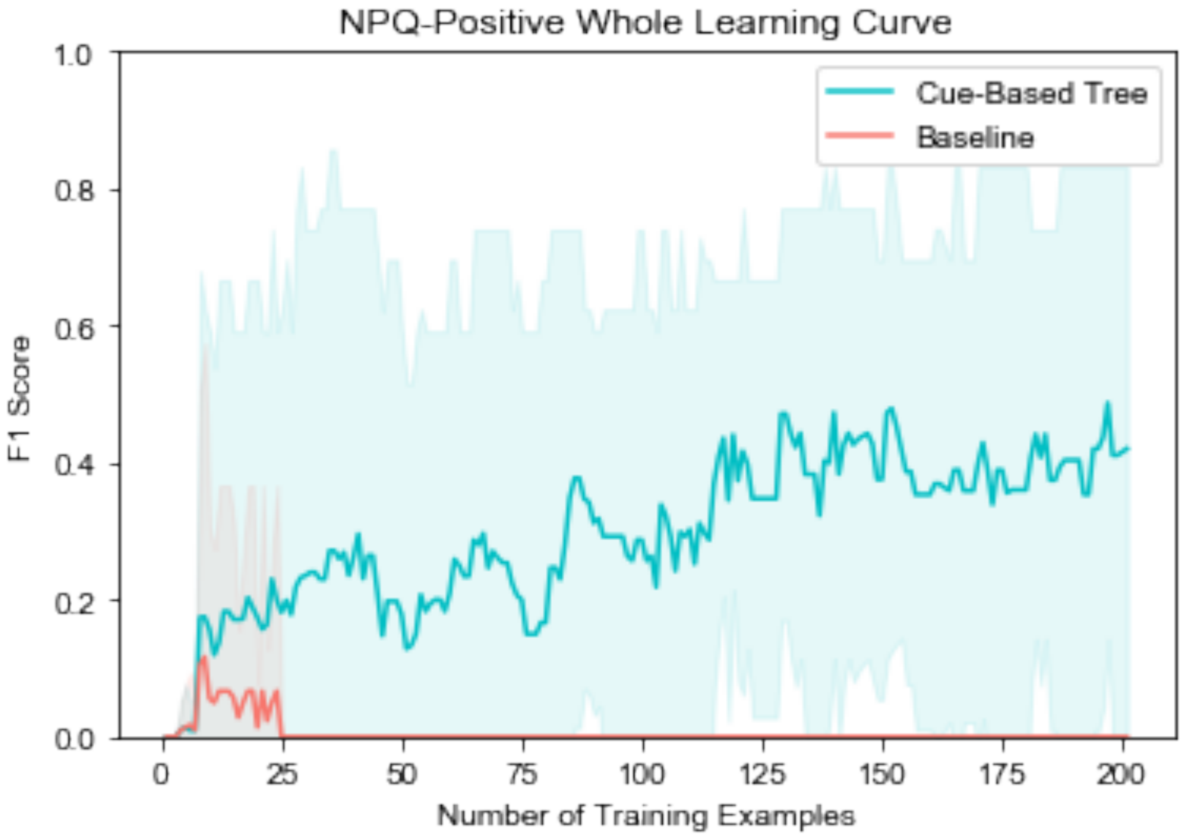


Figure 31. The average F1 score for class NPQ as a function of the number of training examples in the baseline and cue-based models. The colored shades show the 95% confidence intervals.

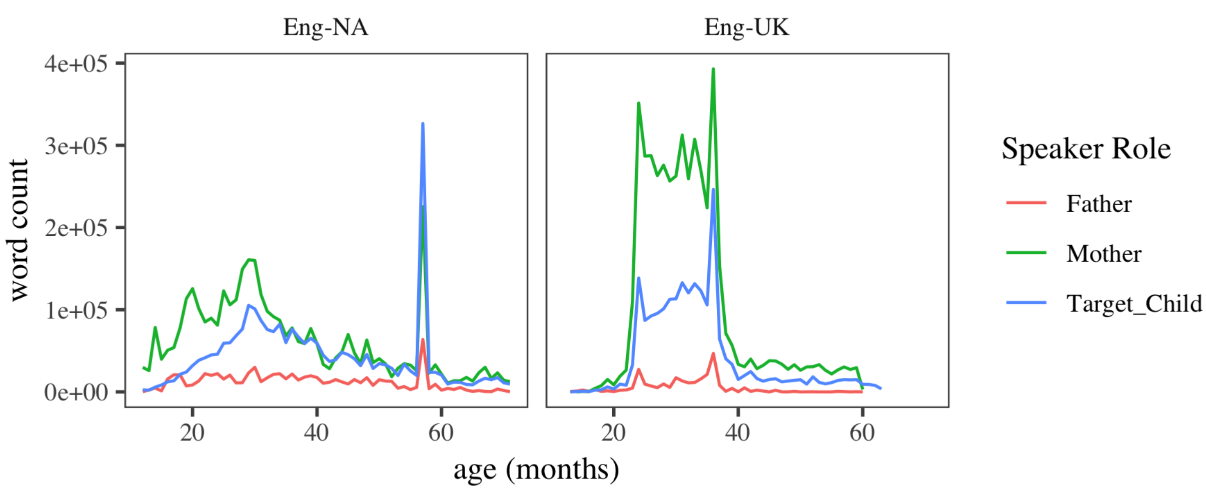


Figure 32. Frequency for all the words in the North America and UK corpora of CHILDES.

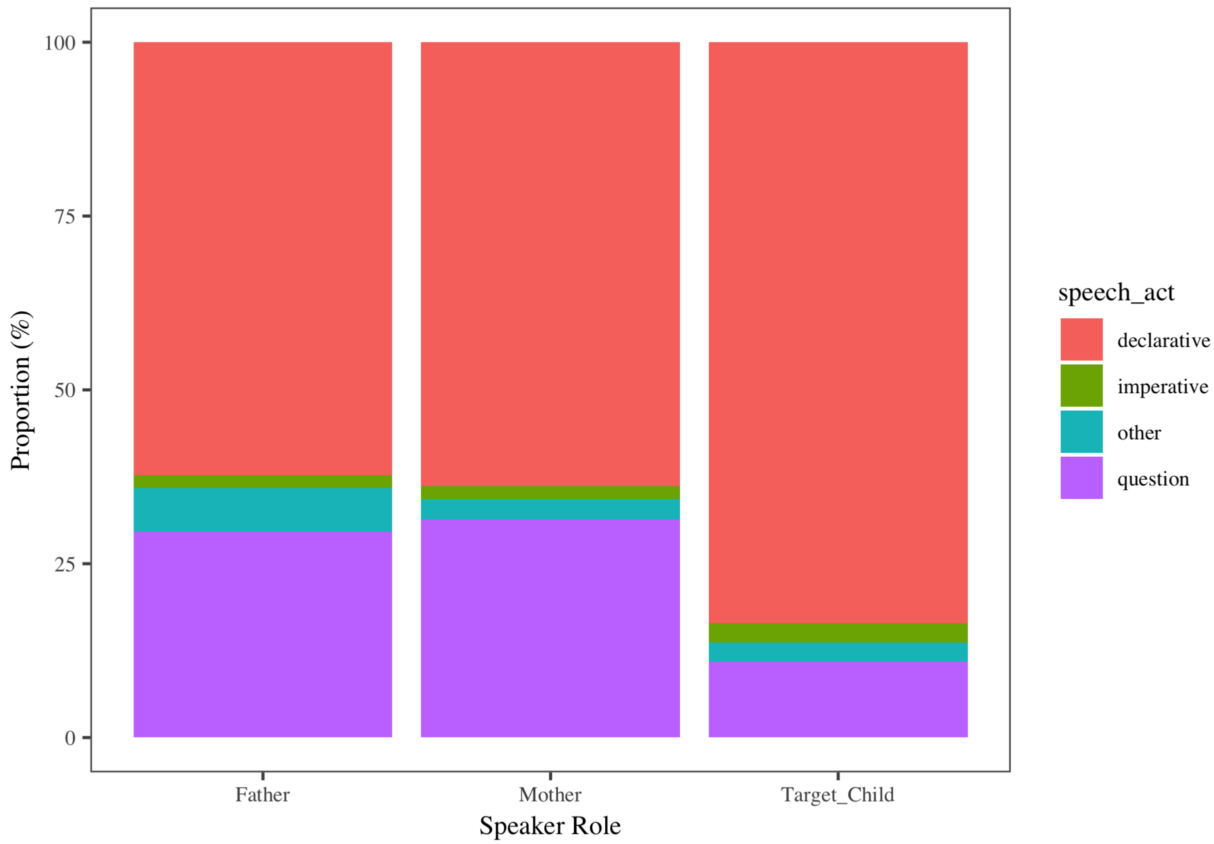


Figure 33. The proportion of declaratives and questions in children's and parents' utterances.

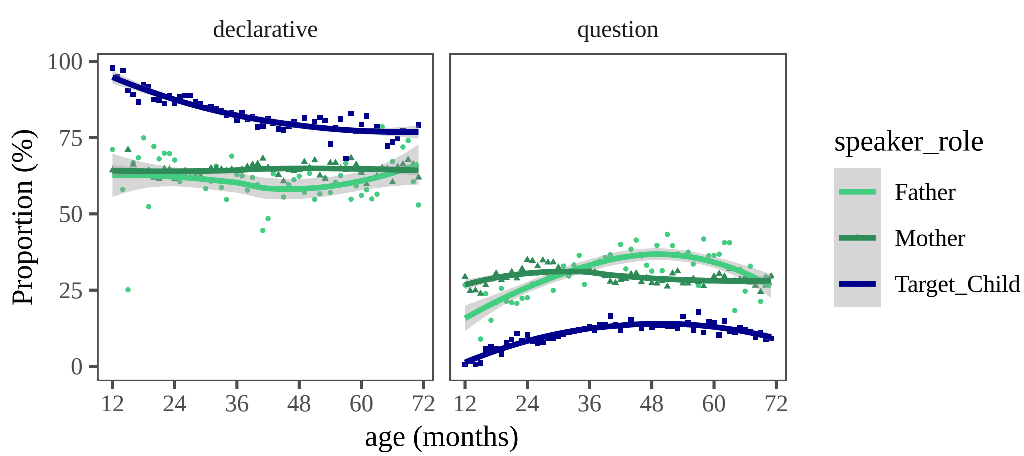


Figure 34. Proportion of declaratives to questions in parent-child interactions by age.

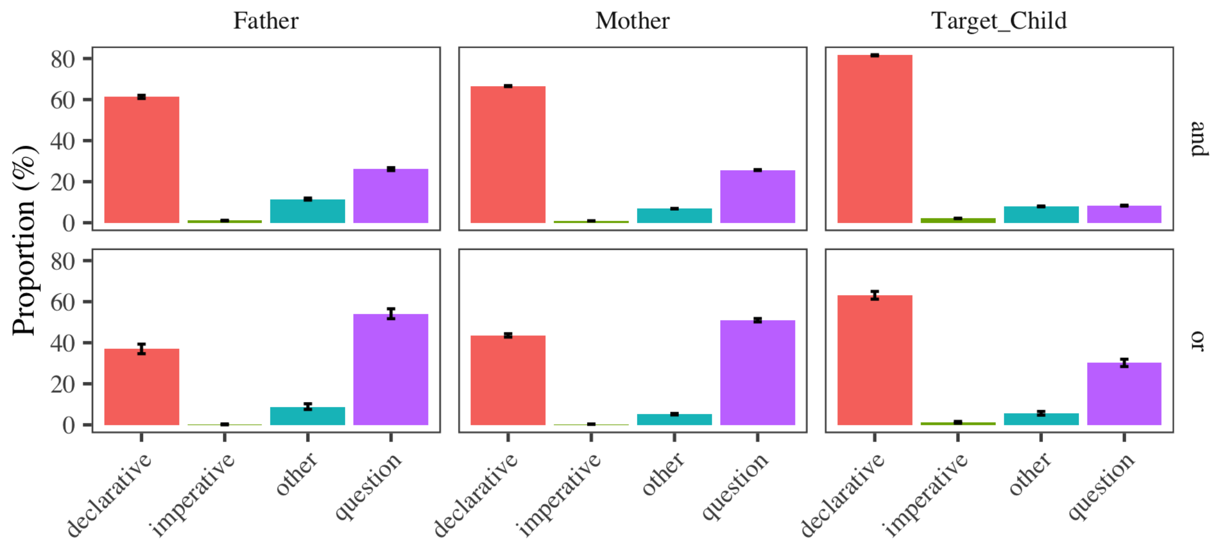


Figure 35. The proportion of *and* and *or* in different utterance types in the speech of parents and children.

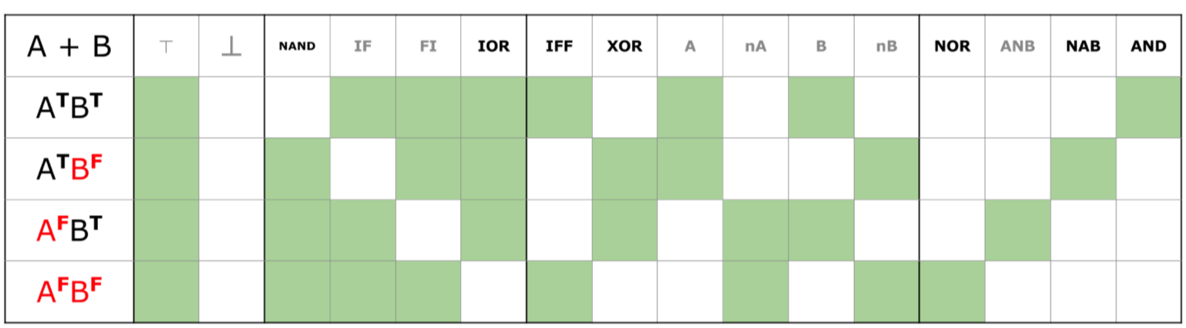


Figure 36. The truth table for the 16 binary logical connectives. The rows represent the set of situations where zero, one, or both propositions are true. The columns represent the 16 possible connectives and their truth conditions. Green cells represent true situations.

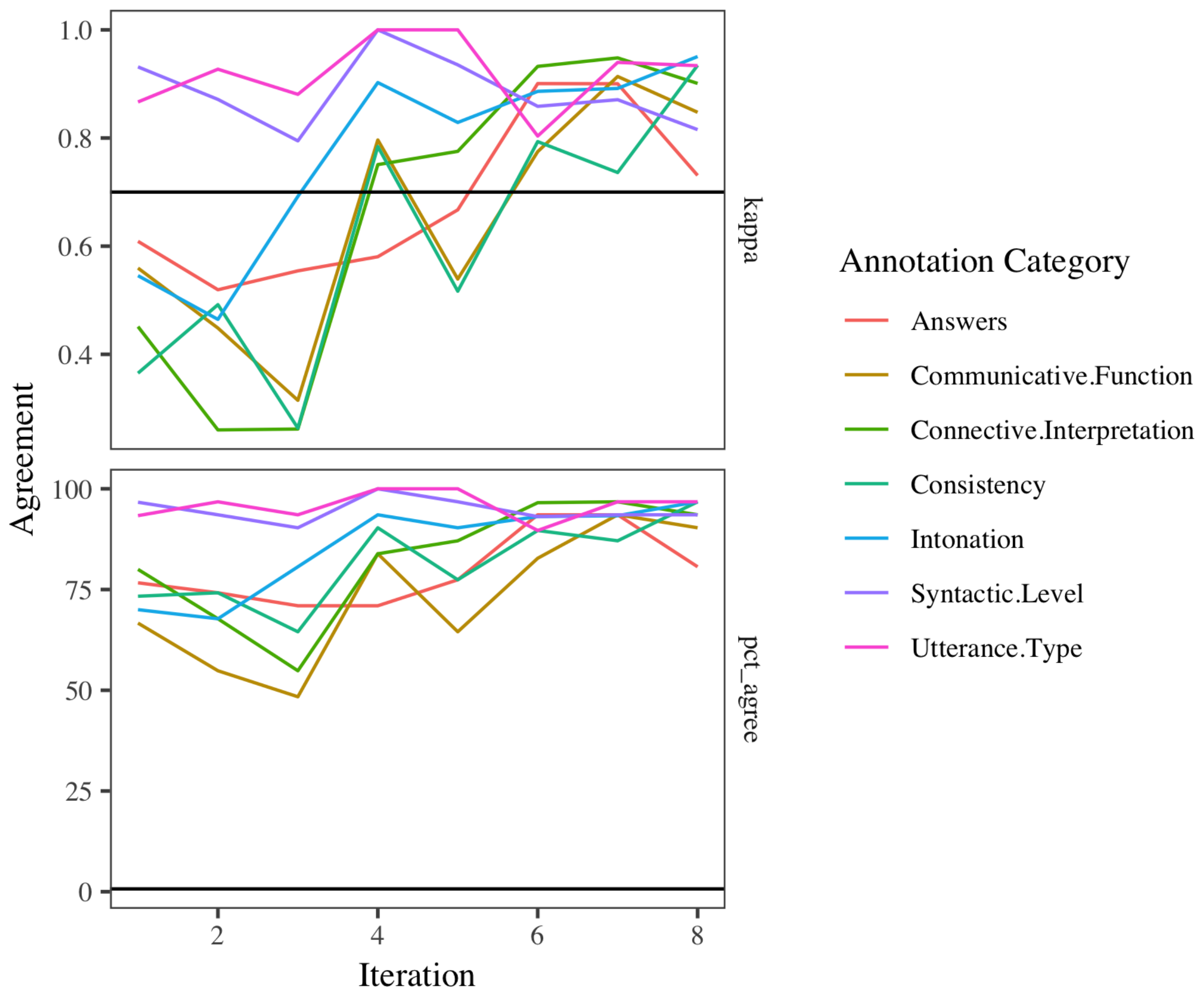


Figure 37. Inter-annotator agreement for disjunction examples.

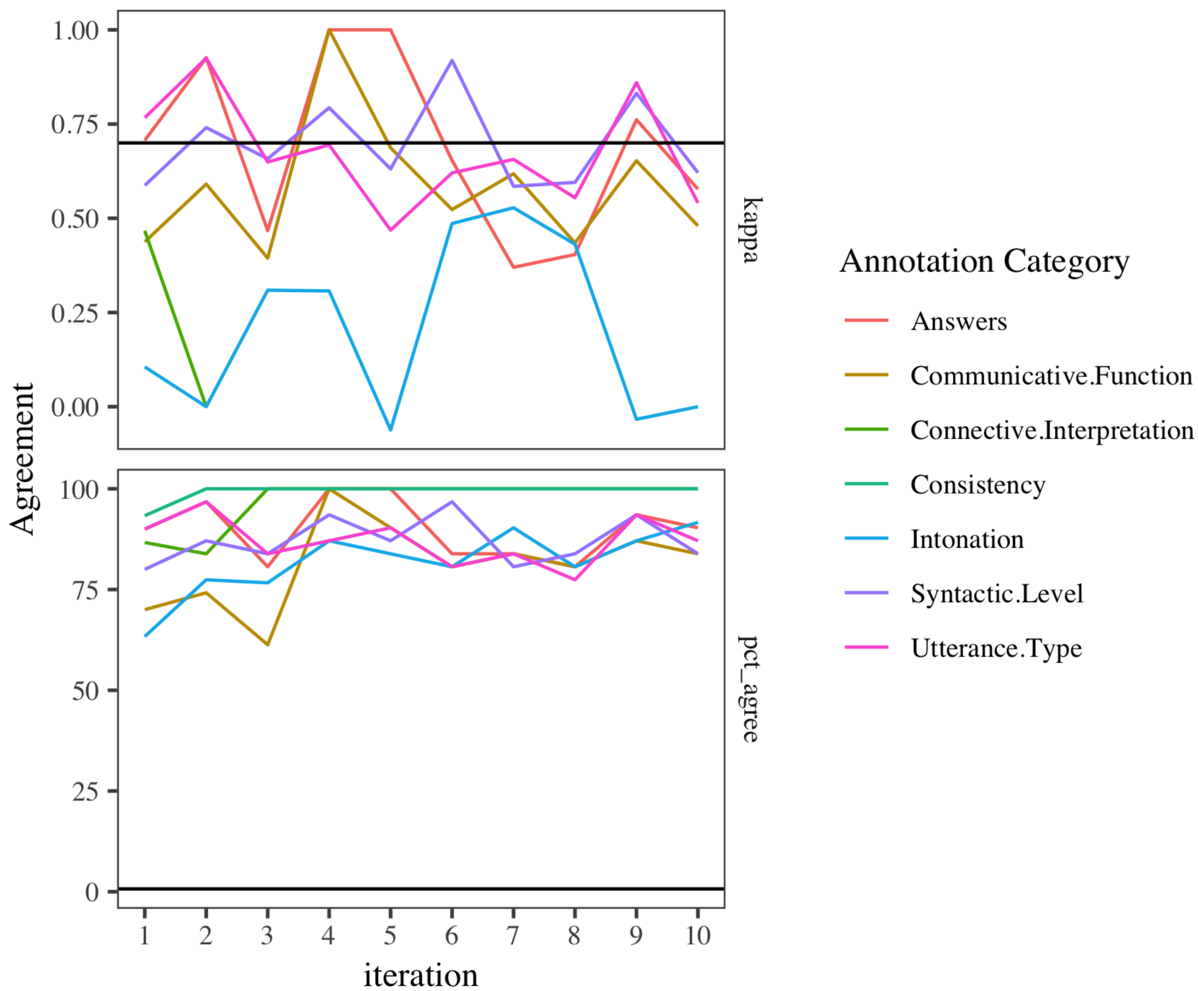


Figure 38. Inter-annotator agreement for conjunction examples.